AD-A202 507

WHOI-88-2

מית הדה תקר

Woods Hole Oceanographic Institution

DTIC
S ELECTE
DEC 0 6 1988
D

WOODS HOLE OCEANOGRAPHIC INSTITUTION
1930

Time Delay Estimation in Stationary and Non-Stationary Environments

by

Mordechai Segal and Ehud Weinstein

July 1988

Technical Report

88 12 5 023

Time Delay Estimation in Stationary and Non-Stationary Environments

by

Mordechai Segal and Ehud Weinstein

Woods Hole Oceanographic Institution
Woods Hole, Massachusetts 02543

July 1988

Technical Report

Approved for Distribution:

Albert J. Williams III, Chairman
Department of Ocean Engineering

ABSTRACT

We develop computationally efficient iterative algorithms for finding the Maximum Likelihood estimates of the delay and spectral parameters of a noise-like Gaussian signal radiated from a common point source and observed by two or more spatially separated receivers. We first consider the stationary case in which the source is stationary (not moving) and the observed signals are modeled as wide sense stationary processes. We then extend the scope by considering a non-stationary (moving) source radiating a possible non-stationary stochastic signal. In that context, we address the practical problem of estimation given discrete-time observations. We also present efficient methods for calculating the log-likelihood gradient (score), the Hessian, and the Fisher's information matrix under stationary and non-stationary conditions.

I. INTRODUCTION

Estimation of the time delay between signals radiated from a common point source and observed at two or more spatially separated receivers is a problem of considerable practical interest (e.g. see [1]). Assuming that the source signal and the additive receiver noises are mutually independent wide sense stationary Gaussian processes with known spectra, and that the observation interval is long compared with the correlation time (inverse bandwidth) of the signal and the noises, the Maximum Likelihood (ML) estimate of the pairwise differential delay is obtained by pre-filtering and cross-correlating the received signals, and searching for the peak of the cross-correlator output [2] [3]. Under the stated assumptions, the ML delay estimate is optimal in the sense that it is asymptotically unbiased and its error variance approaches the Cramer-Rao lower bound.

In practice, the signal and noise spectra are not precisely known. One is unlikely to have accurate prior information about signal bandwidth, center frequency, or power level, and the statistical description of the noise field is similarly incomplete. In [4] it has been shown that lack of knowledge of the spectral parameters does not affect the quality (mean square error) of the delay estimate. However, this is true only if the joint ML estimation of the delay and the unknown spectral parameters is carried out, and the estimation errors are made sufficiently small.

Unfortunately, for most cases, the joint ML estimation of the delay and spectral parameters involves a complicated multi-dimensional optimization, and therefore it has not been attempted in practice. A common ad-hoc approach [3] consists of estimating the signal and noise spectra (or, alternatively, the coherence function) prior to the cross-correlation operation. However, this procedure is only sub-optimal, and its inherent accuracy critically depends on the method employed for spectral estimation.

A further complication in the delay estimation problem arises if the signal source is moving relative to the receivers, causing to a time-varying delay. Measurement of the delay derivative, that is the Doppler time-compression, provides important additional information concerning source location

and track. However, the time-varying delay causes the signals observed at different receivers to be jointly non-stationary, even if the signal at each receiver output is stationary. An approximate ML scheme, under the assumption of linearly time-varying delay (constant Doppler) was developed by Knapp and Carter [5]. Basically, it forms the cross-correlation of one receiver output with respect to a time-delayed and time-scaled version of the other receiver output, and obtain the joint ML estimate of the delay and Doppler parameters by maximizing the cross-correlation response. An alternative scheme that does not depend on the constant Doppler assumption, and by-pass the need for the complicated time-scaling operation was developed in [6]. It basically consists of partitioning the observation interval into time segments, obtain the estimate of the average delay at each time segment, and then apply a least-squares procedure to convert these estimates to the estimate of the delay, Doppler and higher order derivatives. If the duration of each time segment is long compared with the correlation time of the signal and the noises, this estimation procedure is nearly optimal in the sense that it yields an asymptotically unbiased minimum variance estimate of the delay parameters. As pointed out before, if there are unknown spectral parameters, they must be included in the estimation process in order to maintain the asymptotic efficiency of the delay estimates, and this may drastically complicate the procedures in [5] and in [6].

If the signal and noise processes are not stationary over the observation interval, the delay estimation problem is significantly more complicated, and most of the analyses (e.g., [2]-[6]) cannot be applied. In a recent paper [7], Stuller derives the log-likelihood function under the conditions of time-varying delay and possibly non-stationary source signal contaminated by additive white Gaussian noises. Basically, it involves the solution to an integral equation for each value of the unknown parameters (delay, Doppler, amplitude attenuations, spectral parameters) in the a-priori domain. Therefore, the full multi-dimensional search for the ML parameter estimates tends to be computationally complex and time consuming. In a subsequent paper [8], Namazi and Stuller take a different approach; they view the two-channel time delay estimation as a demodulation problem, where the time varying delay is modeled as a realization from a stochastic process. They propose iterative

algorithms, based on the ML, Maximum-A-Posteriori (MAP), and Minimum-Mean-Square-Error (MMSE) criteria, for simultaneous estimation of the source signal and the delay process. Each iteration cycle requires the solution of an integral equation to obtain the up-dated estimate of the signal, and an integral equation to obtain the up-dated estimate of the time-varying delay. In addition to being numerically tedius, there is no proof of convergence of these algorithms, and if they converge it is unclear where they converge to. Another limitation of these algorithms is that they assume prior knowledge of the covariance functions of the signal and the delay processes.

A practical problem that has been overlooked in most of the analyses, is the use of sampled data. If the received signals are first sampled and then processed, the cross-correlation function can be calculated only at a discrete set of time-lags, and the delay estimate is subject to the sampling period. A common numerical solution is to apply a polynomial fitting procedure to interpolate between adjacent values of the cross-correlation function, and then maximize with respect to a continuous delay. An alternative approach [9] is to assume that the signal and the noises are strictly bandlimited, and to use sinc(·) functions to interpolate. However, these approaches are ad-hoc, they can be applied only under constant sampling rates, and they do not necessarily improve the quality of the delay estimate.

In this paper we develop computationally efficient algorithms for the joint estimation of the delay and spectral parameters, based on the Estimate-Maximize (EM) method. The proposed algorithms are optimal in the sense that they converge iteratively to a stationary point of the likelihood function, where each iteration increases the likelihood of the estimated parameters.

The organization of the paper is as follows: In section II we consider the stationary case in which the differential delay is assumed to be constant over the observation interval (stationary source), and the signal and noise statistics are assumed to be stationary. In section III we extend the scope by considering non-stationary sources, and possibly non-stationary signals. We also address the problem of discrete-time observations in a natural way by making an essential use of the continuous-time signals propagating through the medium. In section IV we present an efficient method

to compute the log-likelihood gradient (score), the Hessian, and the Fisher's information matrix (FIM) of the underlying stochastic system. These results can be used for efficient implementation of gradient-search algorithms, and to assess the mean square errors of the ML parameter estimates. Finally, in section V we summarize the results.

II. DELAY ESTIMATION - STATIONARY CASE

A. Problem Formulation and Existing Results

Signals radiated from a stationary (non-moving) point source and observed in the presence of additive noise by a pair of spatially separated receivers can be mathematically modeled as

$$z_1(t) = s(t) + v_1(t) \tag{1}$$

$$z_2(t) = \alpha s(t-\tau) + v_2(t) \tag{2}$$

where τ is the time difference of arrival (TDOA) of the signal wavefront between the receivers.

Suppose that $s(t)$, $v_1(t)$, and $v_2(t)$ are sample functions from mutually independent, wide sense stationary (WSS), zero-mean Gaussian processes with the spectral densities $P_S(\omega;\theta)$, $P_{v_1}(\omega;\theta)$, and $P_{v_2}(\omega;\theta)$, respectively. The vector θ represents possibly unknown spectral parameters such as signal bandwidth, center frequency, noise spectral levels, etc.

Given continuous-time observations of $z_1(t)$ and $z_2(t)$ $T_i \leq t \leq T_f$, we want to find the ML estimate of τ. Since the amplitude attenuation α and the spectral parameters θ are also unknown, they must be included in the estimation process. We denote by

$$\xi = \begin{bmatrix} \tau \\ \alpha \\ \theta \end{bmatrix} \tag{3}$$

the vector unknown parameters to be estimated.

Fourier analyzing $z_1(t)$ and $z_2(t)$,

$$Z_i(\omega_n) = \frac{1}{\sqrt{T}} \int_{T_i}^{T_f} z_i(t) \, e^{-j\omega_n t} \, dt \, , \quad \omega_n = \frac{2\pi}{T} \cdot n \tag{4}$$

where $T = T_f - T_i$. Assuming that the observation interval T is long compared with the correlation time (inverse bandwidth) of the signal and the noises (i.e., $WT/2\pi \gg 1$), the $\underline{Z}(\omega_n) = \left[Z_1(\omega_n) \; Z_2(\omega_n) \right]^T$ $n = 1,2,...$ are statistically uncorrelated zero-mean and Gaussian with the covariance matrix

$$P(\omega_n;\xi) = E\{\underline{Z}(\omega_n)\underline{Z}^*(\omega_n)\}$$

$$= \begin{bmatrix} P_S(\omega_n;\theta) + P_{V_1}(\omega_n;\theta) & \alpha e^{j\omega_n \tau} P_S(\omega_n;\theta) \\ \alpha e^{-j\omega_n \tau} P_S(\omega_n;\theta) & \alpha^2 P_S(\omega_n;\theta) + P_{V_2}(\omega_n;\theta) \end{bmatrix} \tag{5}$$

Therefore, the probability density of $\underline{Z}(\omega_n)$ is

$$f[\underline{Z}(\omega_n)] = \frac{1}{\det[\pi P(\omega_n;\xi)]} e^{-\underline{Z}^*(\omega_n)P^{-1}(\omega_n;\xi)\underline{Z}(\omega_n)} \tag{6}$$

and the log-likelihood function is

$$L_Z(\xi) = \sum_n \log f[\underline{Z}(\omega_n)]$$

$$= - \sum_n \{\log \det[\pi P(\omega_n;\xi)] + \underline{Z}^*(\omega_n) \, P^{-1}(\omega_n;\xi)\underline{Z}(\omega_n)\} \tag{7}$$

Substituting (5) into (7) and carrying out the indicated matrix manipulations,

$$L_Z(\xi) = - \sum_n \log \pi\Delta(\omega_n;\alpha,\theta)$$

$$- \sum_n \frac{1}{\Delta(\omega_n;\alpha,\theta)} \{[\alpha^2 P_S(\omega_n;\theta) + P_{V_2}(\omega_n;\theta)]\cdot|Z_1(\omega_n)|^2$$

$$+[P_S(\omega_n;\theta) + P_{V_1}(\omega_n;\theta)]\cdot|Z_2(\omega_n)|^2 - 2\alpha\mathrm{Re}[e^{j\omega_n\tau} \, P_S(\omega_n;\theta)Z_1^*(\omega_n)Z_2(\omega_n)]\} \tag{8}$$

where Re[·] stands for the real part of the bracketed quantity, and

$$\Delta(\omega_n;\alpha,\theta) = [\alpha^2 P_{V_1}(\omega_n;\theta) + P_{V_2}(\omega_n;\theta)] \, P_S(\omega_n;\theta) + P_{V_1}(\omega_n;\theta)P_{V_2}(\omega_n;\theta) \tag{9}$$

The ML estimates of the unknown parameters require the maximization of $L_Z(\xi)$ with respect to ξ, that is

$$\underset{\xi}{\mathrm{Max}} \, L_Z(\xi) =====> \hat{\xi} \tag{10}$$

For most cases, this is a complicated multi-parameter optimization problem and therefore it has not been attempted.

Following the considerations in [3], we observe that the log-likelihood as a function of τ can be rewritten in the form

$$L_Z(\tau) = c + 2\text{Re}\left\{ \sum_n W(\omega_n;\alpha,\theta)\ Z_1^*(\omega_n)\ Z_2(\omega_n)\ e^{j\omega_n\tau} \right\}$$

$$= c + \frac{T}{\pi}\,\text{Re}\left\{ \sum_n W(\omega_n;\alpha,\theta)Z_1^*(\omega_n)\ Z_2(\omega_n)\ e^{j\omega_n\tau}\ \Delta\omega \right\}$$

$$\xrightarrow[WT/2\pi \gg 1]{} c + \frac{T}{\pi}\,\text{Re}\left\{ \int_\omega W(\omega;\alpha,\theta)\ Z_1^*(\omega)Z_2(\omega)e^{j\omega\tau}\ d\omega \right\} \qquad (11)$$

where c is a constant independent of τ, and

$$W(\omega;\alpha,\theta) = \frac{\alpha P_S(\omega;\theta)}{[\alpha^2 P_{V_1}(\omega;\theta)+P_{V_2}(\omega;\theta)]\ P_S(\omega;\theta)+P_{V_1}(\omega;\theta)P_{V_2}(\omega;\theta)} \qquad (12)$$

The integral in (11), that is the inverse Fourier transform of $W(\omega;\alpha,\theta)Z_1^*(\omega)Z_2(\omega)$, is termed the Generalized Cross Correlation (GCC) [3]. Thus, if we have had exact prior knowledge of α and θ, the ML estimate of the delay parameter would be obtained by plotting the real part of the GCC as a function of delay, and search for its peak, that is

$$\underset{\tau}{\text{Max}}\,\text{Re}\left\{ \int_\omega W(\omega;\alpha,\theta)Z_1^*(\omega)Z_2(\omega)e^{j\omega\tau}\ d\omega \right\} ====> \hat{\tau} \qquad (13)$$

As pointed out by Knapp and Carter [3], the ML estimator is identical to the estimation scheme developed by Hannan and Thomson [10] (see also Hamon and Hannan [11]). Other delay estimation techniques ([12]-[13]) have the same format as (13), where the weighting function $W(\omega;\alpha,\theta)$ is chosen to optimize a selected criterion (e.g., signal to noise ratio, detection index, etc.). These methods are expected to outperform the conventional cross-correlation instrumentation ($W(\omega)=1$) by taking full advantage of the spectral details of the signal and the noises.

In practice, we do not have prior knowledge of α, and the spectral description of the signal and the noises is incomplete. Therefore, it has been suggested first to estimate the selected weighting

function, and then use it in (13) (e.g., see [3], [9], [11], [14]). However, this approach is only sub-optimal, and its inherent performance critically depends on the method employed for spectral estimation.

The Cramer-Rao lower bound (CRLB) on the error variance of any unbiased estimator of the delay is given by (e.g., see [15], Chapter 2)

$$VAR(\hat{\tau}) \geq [J^{-1}(\xi)]_{11} \tag{14}$$

where $J(\xi)$ is the Fisher's information matrix (FIM) associated with all the unknown parameters in the problem. In [4] it has been shown that the FIM processes the following block-diagonal form

$$J(\xi) = \begin{bmatrix} J_1 & 0^T \\ 0 & J_2 \end{bmatrix} \tag{15}$$

where

$$
\begin{aligned}
J_1 &= E\{[\partial L_Z(\xi)/\partial\tau]^2\} \\
&= \frac{T}{\pi} \int_0^\infty \frac{\alpha^2\omega^2 P_S^2(\omega;\theta)}{[\alpha^2 P_{V_1}(\omega;\theta)+P_{V_2}(\omega;\theta)]P_S(\omega;\theta)+P_{V_1}(\omega;\theta)P_{V_2}(\omega;\theta)} \, d\omega
\end{aligned} \tag{16}
$$

Therefore,

$$VAR(\hat{\tau}) \geq 1/J_1 \tag{17}$$

Invoking the asymptotic efficiency and lack of bias of the ML estimator, the lower bound in (17) is attainable. However, this is true only if the ML estimation of all the unknown parameters (that is, the solution to (10)) is performed. If we first estimate the spectral weighting function and

then substitute it in (13), we may obtain a biased estimate whose error variance may be significantly larger than the lower bound.

In the next sub-section, we develop a computationally efficient iterative algorithm for finding the joint ML estimate of the delay and spectral parameters without the need to solve the complicated multi-parameter optimation indicated in (10).

B. Development of the Algorithm

In this section we apply the Estimate-Maximize (EM) algorithm [16] to the two-channel time-delay estimation problem. The EM algorithm is an iterative method for finding the ML parameter estimates given incomplete data. It works with a complete data specification, and iterates between estimating the log-likelihood of the complete data using the observed (incomplete) data and the current parameter estimates, and maximizing the estimated log-likelihood function to obtain the new parameter estimates.

If we have had the observation of the source signal s(t) in addition to $z(t) = \left[z_1(t)\ z_2(t) \right]^T$, the estimation problem would be relatively easy. Therefore, we choose as the complete data y(t) the vector

$$y(t) = \begin{bmatrix} z_1(t) \\ z_2(t) \\ s(t) \end{bmatrix} = \begin{bmatrix} z(t) \\ s(t) \end{bmatrix} \quad T_i \leq t \leq T_f \tag{18}$$

Fourier analyzing y(t),

$$\underline{Y}(\omega_n) = \begin{bmatrix} Z_1(\omega_n) \\ Z_2(\omega_n) \\ S(\omega_n) \end{bmatrix} = \begin{bmatrix} \underline{Z}(\omega_n) \\ S(\omega_n) \end{bmatrix} \tag{19}$$

where $Z_i(\omega_n)$ $i = 1,2$ are defined by (4), and

$$S(\omega_n) = \frac{1}{\sqrt{T}} \int_{T_i}^{T_f} s(t) e^{-j\omega_n t} \, dt \quad , \quad \omega_n = \frac{2\pi}{T} n \tag{20}$$

We note that

$$Z_1(\omega_n) = S(\omega_n) + V_1(\omega_n) \tag{21}$$

$$Z_2(\omega_n) = \alpha e^{-j\omega_n \tau} S(\omega_n) + V_2(\omega_n) \tag{22}$$

where $V_i(\omega_n)$ $i = 1,2$ are the Fourier coefficients of $v_i(t)$ $i = 1,2$. We further note that given $S(\omega_n)$, $Z_1(\omega_n)$ and $Z_2(\omega_n)$ are conditionally independent. Therefore, under the large WT assumption, the log-likelihood of the complete data is

$$L_Y(\xi) = \sum_n \log f[\underline{Y}(\omega_n)]$$

$$= \sum_n \{\log f[S(\omega_n)] + \log f[Z_1(\omega_n)/S(\omega_n)] + \log f[Z_2(\omega_n)/S(\omega_n)]\} \tag{23}$$

where

$$\log f[S(\omega_n)] = -\log \pi P_S(\omega_n;\theta) - |S(\omega_n)|^2/P_S(\omega_n;\theta) \tag{24}$$

$$\log f[Z_1(\omega_n)/S(\omega_n)] = -\log \pi P_{V_1}(\omega_n;\theta) - \frac{|Z_1(\omega_n)-S(\omega_n)|^2}{P_{V_1}(\omega_n;\theta)} \qquad (25)$$

$$\log f[Z_2(\omega_n)/S(\omega_n)] = -\log \pi P_{V_2}(\omega_n;\theta) - \frac{|Z_2(\omega_n)-\alpha e^{-j\omega_n\tau}S(\omega_n)|^2}{P_{V_2}(\omega_n;\theta)} \qquad (26)$$

To simplify the exposition, we suppose that the noise spectra $P_{V_i}(\omega;\theta) = P_{V_i}(\omega)$ $i = 1,2$ are perfectly known. The vector θ affects only the spectral density of the signal. In that case, substituting (24) - (26) into (23) and following straightforward algebraic manipulations, we obtain

$$L_Y(\xi) = c + 2\alpha \text{Re}\left\{ \sum_n e^{-j\omega_n\tau}S(\omega_n) Z_2^*(\omega_n)/P_{V_2}(\omega_n) \right\}$$

$$- \alpha^2 \sum_n \frac{|S(\omega_n)|^2}{P_{V_2}(\omega_n)} - \sum_n [\log P_S(\omega_n;\theta) + \frac{|S(\omega_n)|^2}{P_S(\omega_n;\theta)}] \qquad (27)$$

where c is a constant independent of ξ.

We are now ready to apply the EM algorithm. Denote by

$$\hat{\xi}^{(\ell)} = \begin{bmatrix} \hat{\tau}^{(\ell)} \\ \hat{\alpha}^{(\ell)} \\ \hat{\theta}^{(\ell)} \end{bmatrix} \qquad (28)$$

the current estimate of ξ after ℓ iterations of the algorithm. Then, the next iteration cycle is specified in two steps as follows:

E-step: Compute

$$Q(\xi, \hat{\xi}^{(\ell)}) = E_{\hat{\xi}^{(\ell)}} \{L_Y(\xi)/\underline{Z}(\omega_1), \underline{Z}(\omega_2), \ldots\} \tag{29}$$

M-step:

$$\underset{\xi}{Max} \, Q(\xi, \hat{\xi}^{(\ell)}) \rightarrow \hat{\xi}^{(\ell+1)} \tag{30}$$

where $E_{\hat{\xi}^{(\ell)}} \{\cdot/\underline{Z}(\omega_1), \underline{Z}(\omega_2), \ldots\}$ denotes the conditional expectation given the observed data $\underline{Z}(\omega_1)$, $\underline{Z}(\omega_2), \ldots$, evaluated using the current parameter estimate $\hat{\xi}^{(\ell)}$.

Under the continuity of $Q(\xi, \xi')$ in both ξ and ξ' (see [17]), the algorithm converges to a stationary point of $L_Z(\xi)$, the log-likelihood of the observed data, where each iteration cycle increases the likelihood of the estimated parameters. Of course, as in all "hill climbing" algorithms, the convergence point may not be the global maximum of the objective function, and thus several starting points (or, alternatively, a grid search to roughly locate the global maximum) may be needed.

Now, substituting (27) into (29)

$$Q(\xi, \hat{\xi}^{(\ell)}) = c + G^{(\ell)}(\tau, \alpha) + H^{(\ell)}(\theta) \tag{31}$$

where

$$G^{(\ell)}(\tau, \alpha) = 2\alpha Re \left\{ \sum_n e^{-j\omega_n \tau} S(\omega_n)^{(\ell)} Z_2^*(\omega_n)/P_{v_2}(\omega_n) \right\} - \alpha^2 \sum_n \widehat{|S(\omega_n)|^2}^{(\ell)}/P_{v_2}(\omega_n) \tag{32}$$

and

$$H^{(\ell)}(\theta) = - \sum_n \left[\log P_S(\omega_n; \theta) + \frac{\widehat{|S(\omega_n)|^2}^{(\ell)}}{P_S(\omega_n; \theta)} \right] \tag{33}$$

where

$$\hat{S}(\omega_n)^{(\ell)} = E_{\hat{\xi}(\ell)} \{S(\omega_n)/\underline{Z}(\omega_1), \underline{Z}(\omega_2),...\} \tag{34}$$

and

$$\widehat{|S(\omega_n)|^{2(\ell)}} = E_{\hat{\xi}(\ell)} \{|S(\omega_n)|^2/\underline{Z}(\omega_1),\underline{Z}(\omega_2),...\}$$

$$= |\hat{S}(\omega_n)^{(\ell)}|^2 + VAR_{\hat{\xi}(\ell)} \{S(\omega_n)/\underline{Z}(\omega_1), \underline{Z}(\omega_2),...\} \tag{35}$$

We now recall the following well-known theorem (e.g., [18], Chap. 2): If S and Z_1, $Z_2,...Z_N$ are jointly Gaussian zero-mean random variables, and Z_1, $Z_2,...Z_N$ are statistically independent, then

$$E\{S/Z_1,Z_2,...Z_N\} = \sum_{k=1}^{N} E\{S/Z_k\} \tag{36}$$

$$VAR\{S/Z_1,Z_2,...Z_N\} = \sum_{k=1}^{N} VAR\{S/Z_k\} - (N-1)VAR\{S\} \tag{37}$$

Therefore,

$$E_{\hat{\xi}(\ell)} \{S(\omega_n)/\underline{Z}(\omega_1),\underline{Z}(\omega_2),...\}$$

$$= E_{\hat{\xi}(\ell)} \{S(\omega_n)/\underline{Z}(\omega_n)\} + \sum_{k \neq n} E_{\hat{\xi}(\ell)} \{S(\omega_n)/\underline{Z}(\omega_k)\}$$

$$= E_{\hat{\xi}(\ell)} \{S(\omega_n)/\underline{Z}(\omega_n)\} \tag{38}$$

$$VAR_{\hat{\xi}(\ell)} \{S(\omega_n)/\underline{Z}(\omega_1),\underline{Z}(\omega_2),...\}$$

$$= VAR_{\hat{\xi}(\ell)} \{S(\omega_n)/\underline{Z}(\omega_n)\} + \sum_{k \neq n} VAR_{\hat{\xi}(\ell)} \{S(\omega_n)/\underline{Z}(\omega_k)\} - \left(\sum_{k \neq n} \cdot\right) VAR_{\hat{\xi}(\ell)} \{S(\omega_n)\}$$

$$= VAR_{\hat{\xi}(\ell)} \{S(\omega_n)/\underline{Z}(\omega_n)\} \tag{39}$$

where in the transition to the third line of (38) and (39), we invoked the statistical independence

between $S(\omega_n)$ and $\underline{Z}(\omega_k)$ for $k \neq n$. Hence, the conditional expectations in (34) and (35) reduce to:

$$\hat{S}(\omega_n)^{(\ell)} = E_{\hat{\xi}(\ell)} \{S(\omega_n)/\underline{Z}(\omega_n)\} \tag{40}$$

$$\widehat{|S(\omega_n)|^2}^{(\ell)} = E_{\hat{\xi}(\ell)} \{|S(\omega_n)|^2/\underline{Z}(\omega_n)\}$$

$$= |\hat{S}(\omega_n)^{(\ell)}|^2 + VAR_{\hat{\xi}(\ell)} \{S(\omega_n)/\underline{Z}(\omega_n)\} \tag{41}$$

Since $S(\omega_n)$ and $\underline{Z}(\omega_n)$ are jointly Gaussian, then using well-known results (e.g., [18], Chap. 2)

$$E_{\hat{\xi}(\ell)} \{S(\omega_n)/\underline{Z}(\omega_n)\} =$$

$$= E_{\hat{\xi}(\ell)} \{S(\omega_n)\underline{Z}^*(\omega_n)\} \left[E_{\hat{\xi}(\ell)} \{\underline{Z}(\omega_n)\underline{Z}^*(\omega_n)\} \right]^{-1} \underline{Z}(\omega_n) \tag{42}$$

$$VAR_{\hat{\xi}(\ell)} \{S(\omega_n)/Z(\omega_n)\} =$$

$$= VAR_{\hat{\xi}(\ell)} \{S(\omega_n)\} - E_{\hat{\xi}(\ell)} \{S(\omega_n)\underline{Z}^*(\omega_n)\} \left[E_{\hat{\xi}(\ell)} \{\underline{Z}(\omega_n)\underline{Z}^*(\omega_n)\} \right]^{-1} E\{\underline{Z}(\omega_n)S^*(\omega_n)\} \tag{43}$$

In view of (21), (22), and (5)

$$E_{\hat{\xi}(\ell)} \{S(\omega_n)\underline{Z}^*(\omega_n)\} = P_s(\omega_n;\hat{\theta}^{(\ell)}) \left[1 \, \vdots \, \hat{\alpha}(\ell) e^{j\omega_n \hat{\tau}(\ell)} \right] \tag{44}$$

$$E_{\hat{\xi}(\ell)} \{\underline{Z}(\omega_n)\underline{Z}^*(\omega_n)\} = P(\omega_n;\hat{\xi}^{(\ell)}) =$$

$$= \begin{bmatrix} P_s(\omega_n;\hat{\theta}^{(\ell)}) + P_{v_1}(\omega_n) & \hat{\alpha}(\ell) e^{j\omega_n \hat{\tau}(\ell)} \cdot P_s(\omega_n;\hat{\theta}^{(\ell)}) \\ \hat{\alpha}(\ell) e^{-j\omega_n \hat{\tau}(\ell)} \cdot P_s(\omega_n;\hat{\theta}^{(\ell)}) & \hat{\alpha}(\ell)^2 P_s(\omega_n;\hat{\theta}^{(\ell)}) + P_{v_2}(\omega_n) \end{bmatrix} \tag{45}$$

Substituting (44) and (45) into (42) and (43) and following straightforward matrix manipulations,

$$E_{\hat{\xi}(\ell)}\{S(\omega_n)/\underline{Z}(\omega_n)\} =$$

$$= \frac{P_S(\omega_n;\hat{\theta}(\ell))\left[Z_1(\omega_n)+\hat{\alpha}(\ell)e^{j\omega_n\hat{\tau}(\ell)}\dfrac{P_{v_1}(\omega_n)}{P_{v_2}(\omega_n)}Z_2(\omega_n)\right]}{\left[1+\hat{\alpha}(\ell)^2\dfrac{P_{v_1}(\omega_n)}{P_{v_2}(\omega_n)}\right]P_S(\omega_n;\hat{\theta}(\ell))+P_{v_1}(\omega_n)} \tag{46}$$

$$VAR_{\hat{\xi}(\ell)}\{S(\omega_n)/\underline{Z}(\omega_n)\} = \frac{P_{v_1}(\omega_n)P_S(\omega_n;\hat{\theta}(\ell))}{\left[1+\hat{\alpha}(\ell)^2\dfrac{P_{v_1}(\omega_n)}{P_{v_2}(\omega_n)}\right]P_S(\omega_n;\hat{\theta}(\ell))+P_{v_1}(\omega_n)} \tag{47}$$

Now, in view of (31), the maximization in (30) is decomposed into

$$\underset{\tau,\alpha}{Max}\ G^{(\ell)}(\tau,\alpha) \Longrightarrow \hat{\tau}^{(\ell+1)},\ \hat{\alpha}^{(\ell+1)} \tag{48}$$

$$\underset{\theta}{Max}\,H^{(\ell)}(\theta) \Longrightarrow \hat{\theta}^{(\ell+1)} \tag{49}$$

We observe that $G^{(\ell)}(\tau,\alpha)$ (Eq.(32)) is a quadratic function of α, and the α that maximizes $G^{(\ell)}(\tau,\alpha)$ (for any pre-specified τ) is given by

$$\tilde{\alpha}(\tau) = \frac{Re\left\{\displaystyle\sum_n e^{-j\omega_n\tau}\,\hat{S}(\omega_n)^{(\ell)}Z_2^*(\omega_n)/P_{v_2}(\omega_n)\right\}}{\displaystyle\sum_n \widehat{|S(\omega_n)|}^{2(\ell)}/P_{v_2}(\omega_n)} \tag{50}$$

Therefore, the two-parameter maximization in (48) can be carried out in two steps as follows:

$$Max_{\tau}\,G^{(\ell)}(\tau,\tilde{\alpha}(\tau)) \Longrightarrow \hat{\tau}^{(\ell+1)} \tag{51}$$

$$\hat{\alpha}^{(\ell+1)} = \alpha(\hat{\tau}^{(\ell+1)}) \tag{52}$$

where we note that

$$G^{(\ell)}(\tau, \tilde{\alpha}(\tau)) = \frac{\left[\mathrm{Re}\left\{ \sum_n e^{-j\omega_n \tau} \hat{S}(\omega_n)^{(\ell)} Z_2^*(\omega_n)/P_{v_2}(\omega_n) \right\} \right]^2}{\sum_n \widehat{|S(\omega_n)|^{2(\ell)}}/P_{v_2}(\omega_n)} \tag{53}$$

Incorporating all the above considerations, the algorithm assumes the form:

E-step: Compute

$$\hat{S}(\omega_n)^{(\ell)} = \frac{P_S(\omega_n; \hat{\theta}^{(\ell)}) \left[Z_1(\omega_n) + \hat{\alpha}^{(\ell)} e^{j\omega_n \hat{\tau}^{(\ell)}} \dfrac{P_{v_1}(\omega_n)}{P_{v_2}(\omega_n)} Z_2(\omega_n) \right]}{\left[1 + \hat{\alpha}^{(\ell)2} \dfrac{P_{v_1}(\omega_n)}{P_{v_2}(\omega_n)} \right] P_S(\omega_n; \hat{\theta}^{(\ell)}) + P_{v_1}(\omega_n)} \tag{54}$$

$$\widehat{|S(\omega_n)|^{2(\ell)}} = |\hat{S}(\omega_n)^{(\ell)}|^2 + \frac{P_{v_1}(\omega_n) P_S(\omega_n; \hat{\theta}^{(\ell)})}{\left[1 + \hat{\alpha}^{(\ell)2} \dfrac{P_{v_1}(\omega_n)}{P_{v_2}(\omega_n)} \right] P_S(\omega_n; \hat{\theta}^{(\ell)}) + P_{v_1}(\omega_n)} \tag{55}$$

M-step:

$$\underset{\tau}{\mathrm{Max}} \left[\mathrm{Re}\left\{ \sum_n e^{-j\omega_n \tau} \hat{S}(\omega_n)^{(\ell)} Z_2^*(\omega_n)/P_{v_2}(\omega_n) \right\} \right]^2 =\!=\!=\!=\!=\!\Longrightarrow \hat{\tau}^{(\ell+1)} \tag{56}$$

$$\hat{\alpha}^{(\ell+1)} = \frac{\text{Re}\left\{\sum_n e^{-j\omega_n \hat{\tau}^{(\ell+1)}} \hat{S}(\omega_n)^{(\ell)} Z_2^*(\omega_n)/P_{v_2}(\omega_n)\right\}}{\sum_n |S(\omega_n)|^{2(\ell)}/P_{v_2}(\omega_n)} \tag{57}$$

$$\min_\theta \sum_n \left[\log P_S(\omega_n;\theta) + \frac{\widehat{|S(\omega_n)|^{2(\ell)}}}{P_S(\omega_n;\theta)}\right] =====> \hat{\theta}^{(\ell+1)} \tag{58}$$

Perhaps the most striking feature of the algorithm is that it decomposes the spectral parameter optimization from the delay optimization, resulting in a considerable simplification in the computations involved. Since the algorithm is based on the EM method, it must converge to the solution of (10) or, at least, to a stationary point of $L_Z(\xi)$.

We note that Eqs.(56) and (57) are the solution to the ML problem of estimating τ and α assuming that the source signal is known to the observer, where $S(\omega_n)$ and $|S(\omega_n)|^2$ are substituted by their current estimates $\hat{S}(\omega_n)^{(\ell)}$ and $\widehat{|S(\omega_n)|}^{2(\ell)}$, respectively. Similarly, the optimization in (58) yields the solution to the ML problem of estimating the spectral parameters, where the sufficient statistic $|S(\omega_n)|^2$ is substituted by its current estimate. Thus, the algorithm iterates back and forth, using the current parameter estimates to improve the signal estimate and thus to improve the next parameter estimate. The algorithm is illustrated in Figure 1. We note that $\hat{S}(\omega_n)^{(\ell)}$ and $\widehat{|S(\omega_n)|}^{2(\ell)}$ are, in fact, the outputs of the (non-causal) Wiener filter applied to the two-channel data.

As indicated by (15), the quality of the delay estimate is unaffected by lack of exact knowledge of the signal spectral parameters. Therefore, if we are essentially interested in the delay estimation and we are close to the point of convergence, we may consider performing a partial M-step, leaving the spectral estimates at their current value and updating only the estimates of τ and α. By that we may save in computations with only insignificant effect on the rate of convergence of the algorithm and the quality of the resulting delay estimate.

An initial estimate may be obtained by first estimating the spectral parameters of the signal at one receiver output, and then use these estimates to construct the initial estimate of the amplitude attenuation and the delay parameters.

C. Simulation Results

Consider the following example taken from [14]. The observed data are generated by sampling (1)-(2) at a constant rate that is sufficiently high to preserve the spectral structure of the continuous waveforms. The additive noises are assumed to be spectrally white with

$$P_{V_1}(\omega) = P_{V_2}(\omega) = \sigma^2 \quad - \pi/\Delta t \le \omega \le \pi/\Delta t$$

and the source signal is modeled as an all-pole process of order 3, with the spectral density

$$P_S(\omega) = \gamma^2 \cdot |1 + \theta_1 \exp j\omega\Delta t + \theta_2 \exp j2\omega\Delta t + \theta_3 \exp j3\omega\Delta t|^{-2} \quad - \pi/\Delta t \le \omega \le \pi/\Delta t$$

where $\Delta t = T/N$ is the sampling period.

Suppose that γ^2 and σ^2 are known constants. The unknown parameters are $\xi = (\tau, \alpha, \theta)^T$ where $\theta = (\theta_1 \theta_2 \theta_3)^T$.

The algorithm has been tested using the following set of numerical values

N   = 1024

γ^2 = 1

σ^2 = 8.97995 (SNR=1)

$\tau/\Delta t$ = 10.5

α = 1

θ_1 = 1.77, θ_2=1.593, θ_3=0.7047

In Tables 1 and 2 of Figure 2, we have tabulated the outcomes of the algorithm for two initial guesses of the delay $\hat{\tau}^{(0)}/\Delta t$ = 10.0, 11.0. The initial guess of the amplitude and spectral parameters is: $\hat{\alpha}^{(0)}$ = 0.7, $\hat{\theta}_1^{(0)}$ = -1.8, $\hat{\theta}_2^{(0)}$ = 1.7, $\hat{\theta}_3^{(0)}$ = -0.8. For reference, we have also calculated the log-likelihood $L_Z(\xi)$ (Eq.(8)), normalized by the number N of data points, along the iterations. The CRLB (Eq. (16)) on the rms error of the delay estimate is $\sigma(\hat{\tau}/\Delta t)$ = 0.07. We see that after few iterations, the algorithm converges within the CRLB to the ML estimate of the delay and spectral parameters.

In Tables 3 and 4, we have tabulated the results using the same $\hat{\tau}^{(0)}/\Delta t$ = 10.0, 11.0, and the following choice of initial estimates: $\hat{\alpha}^{(0)}$ = 0.7, $\hat{\theta}_1^{(0)}$ = -2.2, $\hat{\theta}_2^{(0)}$ = 2.2, $\hat{\theta}_3^{(0)}$ = -1.5. As we see, after few iterations, the algorithm converges within the CRLB to the ML delay estimate, although the amplitude and spectral estimates exhibit large errors (perhaps due to slow converge rate, or the convergence to a stationary point of the likelihood function).

D. Generalization to Multi-Receivers

Consider the generalization of the proposed algorithm to the estimation of vector delays obtained with an array of M spatially distributed receivers. The observed signals are

$$z_i(t) = \alpha_i\, s(t-\tau_i) + v_i(t) \qquad i=1,...(M-1) \tag{59}$$

$$z_M(t) = s(t) + v_M(t) \tag{60}$$

where τ_i's are the TDOA's relative to the M'th receiver.

We suppose that s(t), $v_1(t),...v_M(t)$ are sample functions from mutually independent, WSS zero-mean Gaussian processes with the spectral densisties $P_s(\omega;\theta)$, $P_{v_1}(\omega),...P_{v_M}(\omega)$, respectively. To simplify the exposition, we have assumed that $P_{v_i}(\omega)$ i=1,...M are perfectly known.

Given continuous-time observation of $z(t)=(z_1(t),...z_M(t))^T$ $T_i \leq t \leq T_f$, we want to find the ML

estimate of the vector delays $\tau=(\tau_1,...\tau_{M-1})^T$ jointly with the amplitude attenuations $\alpha=(\alpha_1,...\alpha_{M-1})^T$

and the unknown signal spectral parameters θ. Denote by

$$\xi = \begin{bmatrix} \tau \\ \alpha \\ \theta \end{bmatrix} \tag{61}$$

the vector unknown parameters to be estimated.

Even if θ and α are known, the search for the ML estimation of τ is a relatively complicated

maximization in (M-1) unknowns. If our ultimate objective is to estimate source location parame-

ters (i.e., bearing and range), we can either estimate the vector delays and convert it to bearing and

range estimate, or to maximize the log-likelihood function directly with respect to bearing and

range, see [19][20].

Consider the application of the EM algorithm. In analog with the two-receiver case, let the

complete data be

$$y(t) = \begin{bmatrix} z_1(t) \\ \cdot \\ \cdot \\ \cdot \\ z_M(t) \\ s(t) \end{bmatrix} = \begin{bmatrix} z(t) \\ s(t) \end{bmatrix} \quad T_i \leq t \leq T_f \tag{62}$$

In the frequency domain,

$$\underline{Y}(\omega_n) = \begin{bmatrix} z_1(\omega_n) \\ \cdot \\ \cdot \\ \cdot \\ Z_M(\omega_n) \\ \\ S(\omega_n) \end{bmatrix} = \begin{bmatrix} Z(\omega_n) \\ \\ S(\omega_n) \end{bmatrix} \tag{63}$$

where

$$Z_i(\omega_n) = \alpha_i \, e^{-j\omega_n \tau_i} \, S(\omega_n) + V_i(\omega_n) \tag{64}$$

$$Z_M(\omega_n) = S(\omega_n) + V_M(\omega_n) \tag{65}$$

For large WT, the $\underline{Y}(\omega_n)$ are uncorrelated and thus, in the Gaussian case, statistically independent. Therefore, the log-likelihood of the complete data is

$$L_Y(\xi) = \sum_n \log f[S(\omega_n)] + \sum_n \sum_{i=1}^{M} \log f[Z_i(\omega_n)/S(\omega_n)] \tag{66}$$

where $\log f[S(\omega_n)]$ is given by (24),

$$\log f[Z_i(\omega_n)/S(\omega_n)] = -\log \pi \, P_{V_i}(\omega_n) - \frac{|Z_i(\omega_n)-\alpha_i e^{-j\omega_n \tau_i} S(\omega_n)|^2}{P_{V_i}(\omega_n)} \tag{67}$$

$$\log f[Z_M(\omega_n)/S(\omega_n)] = -\log \pi \, P_{V_M}(\omega_n) - \frac{|Z_M(\omega_n)-S(\omega_n)|^2}{P_{V_M}(\omega_n)} \tag{68}$$

Following the derivation in Section II-B, we obtain the following algorithm:

<u>E-Step</u>: Compute

$$\hat{S}^{(\ell)}(\omega_n) = \frac{P_s(\omega_n;\hat{\theta}^{(\ell)})\left[Z_M(\omega_n) + \sum_{i=1}^{M-1} \hat{\alpha}_i^{(\ell)} e^{j\omega_n \hat{\tau}_i^{(\ell)}} \frac{P_{v_M}(\omega_n)}{P_{v_i}(\omega_n)} Z_i(\omega_n)\right]}{\left[1 + \sum_{i=1}^{M-1} \hat{\alpha}_i^{(\ell)2} \frac{P_{v_M}(\omega_n)}{P_{v_i}(\omega_n)}\right] P_s(\omega_n;\hat{\theta}^{(\ell)}) + P_{v_M}(\omega_n)} \tag{69}$$

$$\widehat{|S(\omega_n)|^2}^{(\ell)} = |\hat{S}^{(\ell)}(\omega_n)|^2 + \frac{P_{v_M}(\omega_n)P_s(\omega_n;\hat{\theta}^{(\ell)})}{\left[1 + \sum_{i=1}^{M-1} \hat{\alpha}_i^{(\ell)2} \frac{P_{v_M}(\omega_n)}{P_{v_i}(\omega_n)}\right] P_s(\omega_n;\hat{\theta}^{(\ell)}) + P_{v_M}(\omega_n)} \tag{70}$$

<u>M-Step</u>:

<u>Delay and amplitude optimization</u>:

For i=1,...M-1

$$\underset{\tau_i}{\text{Max}} \left\{ \text{Re}\left[\sum_n e^{-j\omega_n \tau_i} \hat{S}^{(\ell)}(\omega_n) \overset{*}{Z}_i(\omega_n)/P_{v_i}(\omega_n)\right]\right\}^2 ====> \hat{\tau}_i^{(\ell+1)} \tag{71}$$

$$\hat{\alpha}_i^{(\ell+1)} = \frac{\text{Re}\left\{\sum_n e^{-j\omega_n \hat{\tau}_i^{(\ell+1)}} \hat{S}^{(\ell)}(\omega_n) \overset{*}{Z}_i(\omega_n)/P_{v_i}(\omega_n)\right\}}{\sum_n \widehat{|S(\omega_n)|^2}^{(\ell)}/P_{v_i}(\omega_n)} \tag{72}$$

Spectral optimization:

$$\underset{\theta}{\text{Min}} \sum_n \left[\log P_s(\omega_n;\theta) + \frac{\widehat{|S(\omega_n)|^2}^{(\ell)}}{P_s(\omega_n;\theta)} \right] =====> \hat{\theta}^{(\ell+1)} \tag{73}$$

(as in (58)).

The algorithm is illustrated in Figure 3. Once again we see that the spatial parameter optimization is decoupled from the spectral parameter optimization. However, the most striking feature of the algorithm is that it decomposes the complicated multi-dimensional optimization associated with the (M-1) pairs of delay and amplitude parameters into the (M-1) optimizations with respect to each parameter pair separately. We therefore suggest to substitute the direct ML optimization that requires either the joint maximization of (M-1) weighted cross-correlators, or the separate maximization of all M(M-1)/2 combinations of cross-correlators (see [21]), with the iterative maximization that only employs (M-1) cross-correlators in parallel.

E. Time-Varying Delays

If there is relative motion between source and receivers the observed differential delay is time-varying. Suppose there is only a small change in array-source geometry during the observation interval so that the differential delay is essentially linearly time-varying, that is

$$\tau(t) = \tau + \dot{\tau} \cdot t \tag{74}$$

where τ is the differential delay at t=0, and $\dot{\tau}$ is the differential Doppler time compression.

The observed signals are

$$z_1(t) = s(t) + v_1(t) \tag{75}$$

$$z_2(t) = \alpha \, s(t-\tau(t)) + v_2(t) \tag{76}$$

where, as before, s(t) is modeled as a sample function from a WSS zero-mean Gaussian process with the spectral density $P_s(\omega;\theta)$, $v_1(t)$ and $v_2(t)$ are assumed to be independent spectrally white with known power level of $N_0/2$.

We note that each receiver output is stationary, but they are not jointly stationary. This is why the delay estimation problem in this case is significantly more complicated (see Knapp and Carter [5]).

We want to extend our algorithm to the estimation of the vector parameters

$$\xi = \begin{bmatrix} \tau \\ \dot{\tau} \\ \alpha \\ \theta \end{bmatrix} \tag{77}$$

The complete data are still given by

$$y(t) = \begin{bmatrix} z_1(t) \\ z_2(t) \\ s(t) \end{bmatrix} \quad T_i \leq t \leq T_f \tag{78}$$

Given $s(t)$, $z_1(t)$ and $z_2(t)$ are statistically independent. Therefore, the log-likelihood of $y(t)$ $T_i \leq t \leq T_f$ is given by

$$L_Y(\xi) = \quad \log f[s(t) \ T_i \leq t \leq T_f] + \log f[z_1(t) \ T_i \leq t \leq T_f / s(t) \ T_i \leq t \leq T_f]$$

$$+ \log f[z_2(t) \ T_i \leq t \leq T_f / s(t) \ T_i \leq t \leq T_f] \tag{79}$$

where

$$\log f[z_1(t) \ T_i \leq t \leq T_f / s(t) \ T_i \leq t \leq T_f] = c_1 - \frac{1}{N_0} \int_{T_i}^{T_f} [z_1(t) - s(t)]^2 \, dt \tag{80}$$

$$\log f[z_2(t) \ T_i \leq t \leq T_f / s(t) \ T_i \leq t \leq T_f]$$

$$= c_2 - \frac{1}{N_0} \int_{T_i}^{T_f} [z_2(t) - s(t - \tau - \dot{\tau} \cdot t)]^2 \, dt \tag{81}$$

where c_1 and c_2 are independent of ξ, and $\log f[s(t), \ T_i \leq t \leq T_f]$ is the log-likelihood of $s(t) \ T_i \leq t \leq T_f$, which is given, under large WT assumption by

$$\log f[s(t), \ T_i \leq t \leq T_f] = - \sum_n \{ \log \pi \, P_s(\omega_n; \theta) + |S(\omega_n)|^2 / P_s(\omega_n; \theta) \} \tag{82}$$

The E-step of the algorithm requires the conditional expectation of $L_Y(\xi)$ given the observed $z_1(t)$ and $z_2(t)$ at the current parameter estimate $\hat{\xi}^{(\ell)}$. Since $z_1(t)$ and $z_2(t)$ are jointly non-stationary,

this computation seems complicated at first glance. However, since $\hat{\xi}^{(\ell)} = \left[\hat{\tau}^{(\ell)}, \dot{\hat{\tau}}^{(\ell)}, ...\right]^T$ is

given to us, we could take instead the conditional expectation with respect to $z_1(t)$ and

$$z_2^{(\ell)}(t) \overset{\Delta}{=} z_2\left[\frac{1}{1-\dot{\hat{\tau}}^{(\ell)}}\, t\right] \tag{83}$$

Since $z_2(t)$ and $z_2^{(\ell)}(t)$ can be obtained from one another by a time-scale (reversible) operation,

then ignoring end effects it should make no difference.

Now, at the current estimate $\hat{\xi}^{(\ell)}$, $z_1(t)$ and $z_2^{(\ell)}(t)$ are jointly stationary and the required condi-

tional expection is easily carried out.

We observe that the expression in (80) is independent of ξ, the expression in (81) depends only

on α, τ and $\dot{\tau}$, and the expression in (82) depends only on θ. Therefore, the algorithm takes the

form:

E-step: Compute

$$\hat{S}^{(\ell)}(\omega_n) = \frac{P_s(\omega_n; \hat{\theta}^{(\ell)})\left[Z_1(\omega_n) + \hat{\alpha}^{(\ell)}\, e^{j\omega_n \hat{\tau}^{(\ell)}/(1-\dot{\hat{\tau}}^{(\ell)})}\, Z_2^{(\ell)}(\omega_n)\right]}{[1 + \hat{\alpha}^{(\ell)2}]P_s(\omega_n; \hat{\theta}^{(\ell)}) + N_0/2} \tag{84}$$

$$\widehat{|S(\omega_n)|^2}^{(\ell)} = |\hat{S}^{(\ell)}(\omega_n)|^2 + \frac{\dfrac{N_0}{2} P_s(\omega_n; \hat{\theta}^{(\ell)})}{[1 + \hat{\alpha}^{(\ell)2}]P_s(\omega_n; \hat{\theta}^{(\ell)}) + N_0/2} \tag{85}$$

$$\hat{s}^{(\ell)}(t) = {}^{-1}[\hat{S}^{(\ell)}(\omega)] \tag{86}$$

$$\widehat{[s(t)]^2}^{(\ell)} = [\hat{s}^{(\ell)}(t)]^2 + \frac{1}{2\pi}\int_\omega \frac{\dfrac{N_0}{2} P_s(\omega; \hat{\theta}^{(\ell)})}{[1+\hat{\alpha}^{(\ell)2}]P_s(\omega; \hat{\theta}^{(\ell)})+N_0/2}\, d\omega \tag{87}$$

M-step:

Delay, Doppler and amplitude optimization

$$\underset{\tau,\dot{\tau}}{\text{Max}} \left\{ 2 \int_{T_i}^{T_f} z_2(t)\, \hat{s}^{(\ell)}(t-\tau-\dot{\tau}\cdot t)dt - \int_{T_i}^{T_f} [\hat{s}^{(\ell)}(t-\tau-\dot{\tau}\cdot t)]^2\, dt \right\}^2 =====> \hat{\tau}^{(\ell+1)}, \hat{\dot{\tau}}^{(\ell+1)} \qquad (88)$$

$$\hat{\alpha}^{(\ell+1)} = \frac{\int_{T_i}^{T_f} z_2(t)\, \hat{s}^{(\ell)}(t-\hat{\tau}^{(\ell+1)}-\hat{\dot{\tau}}^{(\ell+1)}\cdot t)\, dt}{\int_{T_i}^{T_f} \widehat{[s(t-\hat{\tau}^{(\ell+1)}-\hat{\dot{\tau}}^{(\ell+1)})]^2}^{(\ell+1)}} \qquad (89)$$

Spectral optimization

$$\underset{\theta}{\text{Min}} \sum_n \left[\log P_s(\omega_n;\theta) + \frac{\widehat{|S(\omega_n)|^{2(\ell)}}}{P_s(\omega_n;\theta)} \right] ==> \hat{\theta}^{(\ell+1)} \qquad (90)$$

where $Z_2^{(\ell)}(\omega_n)$ is the Fourier coefficient of $z_2^{(\ell)}(t)$ at frequency ω_n and $^{-1}(\cdot)$ denotes the inverse Fourier transform operation.

The algorithm is illustrated in Figure 4. We note that $\hat{s}^{(\ell)}(t)$ is the time-domain output of the Wiener filter.

We observe that the E-step of the algorithm is essentially unaffected by the non-stationary introduced by the time-varying delay.

The spectral parameter optimization is identical to the stationary case and it is carried out separately.

The joint optimization of the delay and Doppler basically consists of correlating $z_2(t)$ with time-scaled and shifted version of the Wiener filter output $\hat{s}^{(\ell)}(t)$.

We point out once again that the algorithm is guaranteed to converge monotonically to a stationary point of the log-likelihood function.

The extension to the M-receiver case is straightforward. The algorithm decomposes the complicated multi-dimensional optimization associated with the (M-1) triples of delay, Doppler and amplitude parameters into (M-1) optimizations with respect to each parameter triple separately. Thus, the complexity of the algorithm is only linearly affected by the number of unknown delay and Doppler parameters.

III. DELAY ESTIMATION - THE NON-STATIONARY CASE.

In this section, we extend the scope by postulating a moving source travelling along some unknown trajectory and radiating a possibly non-stationary noise-like stochastic signal.

A. Problem Formulation

The continuous-time waveforms observed at the receiver outputs are

$$z_1(t) = s(t) + v_1(t) \tag{91}$$

$$z_2(t) = \alpha s[t-\tau(t)] + v_2(t) \tag{92}$$

Let the signal s(t) be modeled using the following continuous-time linear dynamic stochastic state equation

$$\dot{x}(t) = F(t;\theta)x(t) + G(t)w(t) \quad -\infty < t < \infty \tag{93}$$

$$s(t) = h^T x(t) \tag{94}$$

where $x(t)$ is the $q \times 1$ state vector, $w(t)$ is the $r \times 1$ vector independent normalized white Gaussian noise, and θ is the $p \times 1$ vector unknown system (spectral) parameters. This model covers a wide range of signals including continuous-time all-pole and zero-pole processes, transient phenomena, etc.

As suggested in [4] [22], the differential time-varying delay $\tau(t)$ is parametrized using Taylor series expansion

$$\tau(t) = \tau_0 + \tau_1 \cdot t + \tau_2 \cdot t^2 + \dots + \tag{95}$$

which identifies the coefficients τ_j as

$$\tau_j = \frac{1}{j!} \left. \frac{d^j \tau(t)}{dt^j} \right|_{t=0} \tag{96}$$

The τ_j's have direct physical meaning: τ_0 is the differential delay at $t=o$, τ_1 is the differential Doppler time compression at $t=0$. Higher order coefficients correspond to successive Doppler derivatives, all evaluated at $t=0$. If the observation time is short, so that $\tau(t)$ varies essentially linearly, the differential Doppler is constant and the series terminates after two terms. By carrying more terms in the series one allows either longer observation periods or more complex source manuevers.

For practical reasons, the actual data are generated by sampling the time functions in (91) and (92). The resulting discrete-time signals satisfy the following model

$$z_{1k} = s(t_k) + v_{1k} \tag{97}$$

$$z_{2k} = \alpha s[t_k - \tau(t_k)] + v_{2k} \tag{98}$$

We suppose that v_{1k} and v_{2k} are statistically independent zero-mean white Gaussian sequences with unknown spectral levels of $\sigma_1{}^2$ and $\sigma_2{}^2$, respectively.

Given the discrete time observations

$$z_k = \begin{bmatrix} z_{1k} \\ z_{2k} \end{bmatrix} \qquad k = 1,2,...N \tag{99}$$

we want to find the ML estimate of the vector parameters

$$\xi = \begin{bmatrix} r \\ \alpha \\ \sigma_1{}^2 \\ \sigma_2{}^2 \\ \theta \end{bmatrix} \tag{100}$$

where $r = (r_1, r_2...)^T$ are the delay parameters.

The direct ML problem is significantly more complicated because of the source motion, the non-stationary of the signal, and the continuous-discrete nature of the problem.

B. Development of the Algorithm

We want to apply the EM algorithm to solve the problem. Motivated by the considerations in [23], we choose as the complete data the continuous-time state $x(t)$ $T_i \le t \le T_f$ jointly with the observations $z_1, z_2,...z_N$. The log-likelihood of the complete data is therefore given by

$$L_Y(\xi) = L_\chi(\theta) + \sum_{k=1}^{N} \log f(z_{1k}/x(t) \ T_i \le t \le T_f)$$

$$+ \sum_{k=1}^{N} \log f(z_{2k}/x(t) \ T_i \le t \le T_f) \tag{101}$$

where

$$\log f(z_{1k}/x(t) \ T_i \le t \le T_f) =$$

$$= -\frac{1}{2} \log(2\pi\sigma_1^2) \qquad - \frac{1}{2\sigma_1^2} [z_{1k} - s(t_k)]^2 \tag{102}$$

$$\log f(z_{2k}/x(t) \ T_i \le t \le T_f) = -\frac{1}{2} \log(2\pi\sigma_2^2) - \frac{1}{2\sigma_2^2} [z_{2k} - \alpha s[t_k - \tau(t_k)]]^2 \tag{103}$$

and $L_\chi(\theta)$ is the log-likelihood of the continuous state $x(t)$ $T_i \le t \le T_f$ given by (e.g., see [24][25])

$$L_\chi(\theta) = c + \int_{T_i}^{T_f} x(t)^T F(t;\theta)^T \left[G(t)G(t)^T \right]^\# dx(t)$$

$$- \frac{1}{2} \int_{T_i}^{T_f} \text{Tr} \ (F(t;\theta)^T \left[G(t)G(t)^T \right]^\# F(t;\theta)x(t) \ x^T(t)\} dt \tag{104}$$

where c is a constant independent of θ. The symbols $(\cdot)^\#$ and $\text{Tr}(\cdot)$ denote the pseudo-inverse and the trace operations, respectively. Substituting (102)-(104) into (101) and taking the conditional expectation given $z_1, z_2, ... z_N$ at a parameter value $\hat{\xi}^{(\ell)}$,

$$Q(\xi;\xi^{(\ell)}) \triangleq E_{\hat{\xi}^{(\ell)}} \ \{L_Y(\xi)/z_1, z_N\}$$

$$= c_1 + G^{(\ell)}(\tau,\alpha,\sigma_1^2,\sigma_2^2) + H^{(\ell)}(\theta) \tag{105}$$

where c_1 is a constant independent of ξ,

$$G^{(\ell)}(\tau,\alpha,\sigma_1{}^2,\sigma_2{}^2) =$$

$$= -\frac{N}{2}\log\sigma_1{}^2 - \frac{1}{2\sigma_1{}^2}\sum_{k=1}^{N}[z_{1k}{}^2 - 2z_{1k}\,\hat{s}^{(\ell)}(t_k) + \widehat{s^2(t_k)}^{(\ell)}]$$

$$-\frac{N}{2}\log\sigma_2{}^2 - \frac{1}{2\sigma_2{}^2}\sum_{k=1}^{N}[z_{2k}{}^2 - 2\alpha z_{2k}\,\hat{s}^{(\ell)}(t_k;\tau) + \alpha^2\widehat{s^2(t_k;\tau)}^{(\ell)}] \qquad (106)$$

$$H^{(\ell)}(\theta) = \int_{T_i}^{T_f} x(t)^T F(t;\theta)^T\left[G(t)G(t)^T\right]^{\#} dx(t)^{(\ell)}$$

$$-\frac{1}{2}\int_{T_i}^{T_f}\mathrm{Tr}\,\{F(t;\theta)^T\left[G(t)G(t)^T\right]^{\#}F(t;\theta)\widehat{x(t)x(t)^T}^{(\ell)}\}dt \qquad (107)$$

where we have denoted for convenience $s(t;\tau)\overset{\Delta}{=}s[t-\tau(t)]$.

In view of (29), (30) and (105), the algorithm assumes the form:

E-Step: Compute

$$G^{(\ell)}(\tau,\alpha.\sigma_1{}^2,\sigma_2{}^2),\quad H^{(\ell)}(\theta) \qquad (108)$$

M-Step:

$$\underset{\tau,\alpha,\sigma_1{}^2,\sigma_2{}^2}{\mathrm{Max}}\ G^{(\ell)}(\tau,\alpha,\sigma_1{}^2,\sigma_2{}^2) =======> \quad \begin{array}{l}\hat{\tau}^{(\ell+1)},\ \hat{\alpha}^{(\ell+1)}\\ \hat{\sigma}_1{}^{2(\ell+1)},\ \hat{\sigma}_2{}^{2(\ell+1)}\end{array} \qquad (109)$$

$$\underset{\theta}{\mathrm{Max}}\ H^{(\ell)}(\theta) =======> \theta^{(\ell+1)} \qquad (110)$$

As in the stationary case, we observe that the spectral parameter optimization (Eq.(110)) is decoupled from the optimization with respect to the other parameters. Now, for a pre-specified τ, the

α, $\sigma_1{}^2$ and $\sigma_2{}^2$ that maximize $G^{(\ell)}(\cdot)$ can easily be found by differentiation. Therefore, following straightforward algebraic manipulations, we obtain the following algorithm:

E-Step: Compute

$$\hat{x}^{(\ell)}(t) = E_{\hat{\xi}(\ell)} \ \{x(t)/z_1, z_N\} \tag{111}$$

$$\widehat{x(t)x(t)^T}^{(\ell)} = E_{\hat{\xi}(\ell)} \ [x(t)x(t)^T/z_1,...z_N] \tag{112}$$

$$\widehat{\int_{T_i}^{T_f}(\cdot)dx(t)}^{(\ell)} = E_{\hat{\xi}(\ell)} \ \{\int_{T_i}^{T_f} x(t)^T F(t;\theta)^T [G(t)G(t)^T]^{\#} dx(t)/z_1,...z_N\} \tag{113}$$

M-Step:

$$\underset{\tau}{\text{Max}} \ \frac{[\sum_{k=1}^{N} z_{2k} \ \hat{s}^{(\ell)}(t_k;\tau)]^2}{\sum_{k=1}^{N} \widehat{s^2(t_k;\tau)}^{(\ell)}} =====> \hat{\tau}^{(\ell+1)} \tag{114}$$

$$\hat{\alpha}^{(\ell+1)} = \frac{\sum_{k=1}^{N} z_{2k} \ \hat{s}^{(\ell)} (t_k;\hat{\tau}^{(\ell+1)})}{\sum_{k=1}^{N} \widehat{s^2(t_k;\hat{\tau}^{(\ell+1)})}^{(\ell)}} \tag{115}$$

$$\hat{\sigma}_1{}^{2(\ell+1)} = \frac{1}{N} \sum_{k=1}^{N} [z_{1k}{}^2 - 2z_{1k} \ \hat{s}^{(\ell)}(t_k) + \widehat{s^2(t_k)}^{(\ell)}] \tag{116}$$

$$\hat{\sigma}_2^2{}^{(\ell+1)} = \frac{1}{N} \sum_{k=1}^{N} [z_{2k}{}^2 - \hat{\alpha}^{(\ell+1)} z_{2k} \hat{s}^{(\ell)}(t_k;\hat{\tau}^{(\ell+1)})] \tag{117}$$

$$\underset{\theta}{\text{Max}} \; H^{(\ell)}(\theta) =======> \hat{\theta}^{(\ell+1)} \tag{118}$$

where we note that the conditional expectations in (112) and (113) are needed for the computation of $H^{(\ell)}(\theta)$ (Eq.(107)). Since $s(t) = h^T x(t)$ (Eq. (96)), then

$$\hat{s}^{(\ell)}(t) = h^T \hat{x}^{(\ell)}(t) \tag{119}$$

$$s^2(t)^{(\ell)} = h^T \widehat{x(t)x(t)}^{T(\ell)} h \tag{120}$$

Using these relations, the conditional expectations in (111) and (112) are sufficient to compute the terms in (114)-(117). The algorithm is illustrated in Figure 5.

Since the algorithm is based on the EM method, it must converge to a stationary point of $L_Z(\xi)$, the log-likelihood of the observed data, where each iteration cycle increases the likelihood of the estimated parameters.

We note that the delay estimates are not subject to the sampling period, since the maximization in (114) is carried out in the underlying continuous-time domain of the propagating signals. Hence, this scheme is particular useful when we want to estimate the delay with resolution that is a fraction of the sampling interval.

If the stochastic system generating the signal $s(t)$ is known (i.e., θ is known) we simply eliminate Eqs. (113) and (118) from the algorithm, and use this a-priori information in the computation of (111) and (112). If the noise power levels σ_1^2 and σ_2^2 are known a-priori, we simply eliminate Eqs. (116) and (117).

The computation of $\hat{x}^{(\ell)}(t)$ and $\widehat{x(t)x(t)}^{T(\ell)}$ $T_i \leq t \leq T_f$ (Eqs. (111) and (112), respectively) can be carried out by first calculating the discrete-time estimates of $x(t)$ and $x(t)x(t)^T$ using the discrete Kalman smoothing equations, and then exploiting the Markovian nature of the underlying stochastic state equation to interpolate between adjacent time instances. Details of this procedure will be presented in the sequel.

The conditional expectation of the stochastic integral in (113) can be expressed in some important cases in terms of the conditional expectation in (112), e.g., in the case of all-pole signals to be described next. Otherwise, it can be approximated via discretization (see the considerations in [24], and the computationally efficient method developed in [26]).

All-Pole Signals

Let the signal $s(t)$ be modeled as the output of an all-pole filter driven by a white Gaussian noise. The equivalent state model is given by (93)-(94), where

$$
F(t;\theta) = \begin{bmatrix} 0 & 1 & 0 & . & 0 \\ . & & & & . \\ . & & & & . \\ . & & & & 0 \\ 0 & & & & 1 \\ \theta_1 & \theta_2 & . & . & \theta_q \end{bmatrix}
\tag{121}
$$

$$
G(t)^T = [0 \ 0 \ - \ - \ - \ 0\sqrt{g}]
\tag{122}
$$

and

$$
h^T = [1 \ 0 \ - \ - \ - \ 0]
\tag{123}
$$

In this setting, the signal $s(t)$ is stationary; however, since we allow the signal source to manuver, the received signals are not stationary and we could not apply the algorithm developed for the stationary case. Also, we do not assume long observation time, and we relate to the continuous-discrete nature of the problem.

Suppose that the gain g in (122) is known, otherwise it could be factored out of the system equation (93) and inserted into the measurement equation (97) as an unknown amplitude scale.

Substituting (121)-(123) into (104), we obtain (see [24][27])

$$L_\chi(\theta) = c + \frac{1}{g} \theta^T \int_{T_i}^{T_f} x(t)dx_q(t) - \frac{1}{2g} \theta^T \left[\int_{T_i}^{T_f} x(t)x(t)^T dt \right] \theta \tag{124}$$

where $\theta = (\theta_1 \; \theta_2 \; ... \theta_q)^T$, and

$$\left[\int_{T_i}^{T_f} x(t)dx_q(t) \right]_i = \begin{cases} - \int_{T_i}^{T_f} \dot{x}_{i+1}(t) \, x_q(t)dt & i=1,2,...(q-1) \\ \\ - g \, (T_f - T_i)/2 & i=q \end{cases} \tag{125}$$

where $x_i(t)$ denotes the ith component of $x(t)$ (note that $x(t)$ is a q-dimensional vector). Therefore,

$$H^{(\ell)}(\theta) \overset{\Delta}{=} E_{\hat{\xi}(\ell)} \{L_\chi(\theta)/z_1, z_2, ... z_N\}$$

$$= \frac{1}{g} \theta^T \int_{T_i}^{T_f} \widehat{x(t)dx_q(t)}^{(\ell)} - \frac{1}{2g} \theta^T \left[\int_{T_i}^{T_f} \widehat{x(t)x(t)^T}^{(\ell)} dt \right] \theta \tag{126}$$

where

$$\left[\int_{T_i}^{T_f} \widehat{x(t)dx_q(t)}^{(\ell)} \right]_i = \begin{cases} - \int_{T_i}^{T_f} \widehat{x_{i+1}(t)x_q(t)}^{(\ell)} dt & i=1,2,...(q-1) \\ \\ \\ - g\,(T_f - T_i)/2 & i=q \end{cases} \qquad (127)$$

In (126) we have ignored the constant c. Observing that $H^{(\ell)}(\theta)$ is a quadratic function of θ, the maximization required in (118) can be solved analytically. We further observe that the computation of $H^{(\ell)}(\theta)$ requires only the computation of $x(t)x(t)^{T(\ell)}$. Therefore, in the case of all-pole signals, the algorithm assumes the form:

E-Step: Compute

$$\hat{x}^{(\ell)}(t) \text{ and } x(t)x(t)^{T(\ell)} \qquad T_i \le t \le T_f \qquad (128)$$

M-Step:

$$- \underset{r}{\text{Max}}\{\text{Eq.(114)}\} ===> \hat{r}^{(\ell+1)} \qquad (129)$$

$$- \text{Compute } \hat{\alpha}^{(\ell+1)} \text{ using Eq.(115)} \qquad (130)$$

$$- \text{Compute } \hat{\sigma}_1^{2(\ell+1)} \text{ using Eq.(116)} \qquad (131)$$

$$- \text{Compute } \hat{\sigma}_2^{2(\ell+1)} \text{ using Eq.(117)} \qquad (132)$$

$$- \text{Compute}$$

$$\hat{\theta}^{(\ell+1)} = \left[\int_{T_i}^{T_f} \widehat{x(t)x(t)^T}^{(\ell)} dt \right]^{-1} \widehat{\int_{T_i}^{T_f} x(t)dx_q(t)}^{(\ell)} \tag{133}$$

We note that only the estimation of the delay parameters involves maximization. We now describe the computation of $\hat{x}^{(\ell)}(t)$ and $\widehat{x(t)x(t)^T}^{(\ell)}$ in details.

Development of the Continuous-Discrete Smoothing Equations.

The computation of $\hat{x}^{(\ell)}(t)$ (Eq. (111)) and $\widehat{x(t)x(t)^T}^{(\ell)}$ (Eq. (112)) is carried out by considering the equivalent discrete model and calculating the discrete-time estimates of the related quantities, and then exploiting the Markovian nature of the underlying stochastic system to interpolate between adjacent time instances. We shall now describe the proposed procedure step by step. For notational convenience we substitute $\hat{\xi}^{(\ell)}$ by ξ.

The Equivalent discrete model

Let the 2N time points t_k, $t_k - \tau(t_k)$ $k=1,2,...N$ be arranged in increased order, and denote by $\tilde{t}_r$ the rth time instance. Then, in view of (97) and (98), the scalar measurements can be represented as

$$z_r = \lambda_r \, h^T x_r + \rho_r \cdot v_r \qquad r=1,2,...2 \cdot N \tag{134}$$

where $x_r \stackrel{\Delta}{=} x(\tilde{t}_r)$, v_r $r=1,2,...$ are statistically independent normalized Gaussian random variables, and

$$\lambda_r = \begin{cases} 1 & z_r \in \Omega_1 = \{z_{11}, z_{21}, ... z_{N1}\} \\ \alpha & z_r \in \Omega_2 = \{z_{12}, z_{22}, ... z_{N2}\} \end{cases} \tag{135}$$

$$\rho_r = \begin{cases} \sigma_1 & z_r \in \Omega_1 \\ \sigma_2 & z_r \in \Omega_2 \end{cases} \tag{136}$$

Now, following the considerations in (e.g., [25] [28]), the x_r $r=1,2,...$ satisfy the following stochastic difference equation:

$$x_{r+1} = \Phi_{r+1}(\theta) x_r + u_{r+1} \tag{137}$$

where $\Phi_r(\theta)$, the discrete-time state transition matrix, is specified by

$$\Phi_r(\theta) = \Phi(\tilde{t}_r, \tilde{t}_{r-1}; \theta) \tag{138}$$

where $\Phi(t,\tau;\theta)$ is the continuous-time state transition matrix, satisfying the differential equation

$$\frac{\partial \Phi(t,\tau;\theta)}{\partial t} = F(t;\theta)\Phi(t,\tau;\theta) \quad t \geq \tau \tag{139}$$

with the initial condition $\Phi(\tau,\tau;\theta) = I$. The u_r $r=1,2,...$ are statistically independent zero-mean and Gaussian with the covariance matrix

$$Q_r(\theta) = Q(\tilde{t}_r, \tilde{t}_{r-1};\theta) \tag{140}$$

where

$$Q(t,\tau;\theta) = \int_\tau^t \Phi(t,s;\theta)G(s)G(s)^T \Phi(t,s;\theta)^T ds \tag{141}$$

Eqs. (137) and (134) specify the equivalent discrete model.

The discrete smoothing equations

Define

$$\mu_{r|k} = E_\xi \{ x_r / z_1, \ldots z_k \} \tag{142}$$

$$P_{r|k} = E_\xi \{ (x_r - \mu_{r|k}) (x_r - \mu_{r|k})^T / z_1, \ldots z_k \} \tag{143}$$

Then, the discrete Kalman filtering equations are (e.g., [18] [25]):

Propagation equations: For r=1,2,...2N

$$\mu_{r|r-1} = \Phi_r \, \mu_{r-1|r-1} \tag{144}$$

$$P_{r|r-1} = \Phi_r \, P_{r-1|r-1} \, \Phi_r^T + Q_r \tag{145}$$

Up-dating equations: For r=1,2,...,2N

$$\mu_{r|r} = \mu_{r|r-1} + k_r \, (z_r - \lambda_r h^T \mu_{r|r-1}) \tag{146}$$

$$P_{r|r} = (I - \lambda_r k_r h^T) \, P_{r|r-1} \tag{147}$$

where k_r is the Kalman gain defined by

$$k_r = \frac{\lambda_r}{\lambda_r^2 h^T P_{r|r-1} \, h + \rho_r^2} \, P_{r|r-1} h \tag{148}$$

To initiate the recursions in (144)-(147) we must select some initial conditions $\mu_{0|0}$ and $P_{0|0}$ (e.g., $\mu_{0|0} = 0$, and $P_{0|0} = I/\epsilon$ for some $\epsilon > 0$).

At this point we need the definition of

$$P_{r-1,r|k} \overset{\Delta}{=} E_\xi \{(x_{r-1} - \mu_{r-1|k}) (x_r - \mu_{r|k})^T / z_1, z_2 ... z_k\} \tag{149}$$

$$S_{r-1} \overset{\Delta}{=} P_{r-1|r-1} \Phi_r^T P_{r|r-1}^{-1} \tag{150}$$

Then the <u>discrete smoothing equations</u> are (e.g., [25])

For r = 2N,2N-1,...1

$$\mu_{r-1|2N} = \mu_{r-1|r-1} + S_{r-1} (\mu_{r|2N} - \Phi_r \mu_{r-1|r-1}) \tag{151}$$

$$P_{r-1|2N} = P_{r-1|r-1} + S_{r-1} (P_{r|2N} - P_{r|r-1}) S_{r-1}^T \tag{152}$$

and (see Shumway [29])

$$P_{r-1,r|2N} = S_{r-2} P_{r-1|r-1} + S_{r-2} (P_{r-1,r|2N} - P_{r-1|r-1} \Phi_r^T) S_{r-1}^T \tag{153}$$

with the initial condition

$$P_{2N-1,2N|2N} = P_{2N-1|2N-1} \Phi_{2N}^T (I - \lambda_{2N} hk_{2N}^T) \tag{154}$$

<u>The interpolation formulae</u>

Invoking the Markovian nature of the underlying stochastic system, for $\tilde{t}_{r-1} \le t \le \tilde{t}_r$

$$\hat{x}(t) \overset{\Delta}{=} E_\xi \{x(t)/z_1, z_2 ... z_N\}$$
$$= A(t) \mu_{r-1|2N} + B(t) \mu_{r|2N} \tag{155}$$

$$\widehat{x(t)x(t)}^T \overset{\Delta}{=} E_\xi\{x(t)x(t)^T/z_1,z_2...z_N\}$$

$$= \hat{x}(t)\hat{x}(t)^T + A(t) P_{r-1|2N} A(t)^T + B(t) P_{r|2N} B(t)^T$$

$$+ A(t) P_{r-1,r|2N} B(t)^T + B(t) P^T_{r-1,r|2N} A(t)^T + C(t) \tag{156}$$

where

$$A(t) = \Phi(t,\tilde{t}_{r-1}) - B(t) \Phi(\tilde{t}_r,\tilde{t}_{r-1}) \tag{157}$$

$$B(t) = Q(t,\tilde{t}_{r-1}) \Phi(\tilde{t}_r,t)^T [\Phi(\tilde{t}_r,t) Q(t,\tilde{t}_{r-1}) \Phi(\tilde{t}_r,t)^T + Q(\tilde{t}_r,t)]^{-1} \tag{158}$$

$$C(t) = Q(t,\tilde{t}_{r-1}) - B(t)\Phi(\tilde{t}_r,t) Q(t,\tilde{t}_{r-1}) \tag{159}$$

where $\Phi(t,\tau)$ and $Q(t,\tau)$ are defined in (139) and (141), respectively. If $\Phi(t,\tau;\theta)$ and $Q(t,\tau;\theta)$ depend on the difference $(t-\tau)$ (i.e., time invariant system) the interpolation operation can be carried out efficiently using the procedure in [26].

Further Considerations

(1)   Motivated by the considerations in the stationary case, near the point of convergence we may leave the spectral parameters at their current values and perform a partial M-step. This way will still retain the monotonic increase of the likelihood function, and the convergence to a stationary point is guaranteed. From computational point of view, it might be preferable to do so, perhaps at the expense of convergence rate reduction.

(2)   An initial estimate of the signal and its spectral parameters may be obtained by applying the EM-type algorithm proposed in [23] to the signal observed in one receiver output (say, the receiver with the higher SNR). Based on these estimates, an initial estimate of the delay parameters can be obtained. Then we can switch to the full-scale EM algorithm.

(3)   In calculating the conditional expectations (111)-(113), we may consider performing the continuous-time interpolation alone (Eqs. (155)-(156)) while retaining the discrete state estimates

at their current values for some iterations. As proved in [30], this modification preserves the basic properties of the EM algorithm, that is the monotonic increase of the likelihood function on each iteration cycle, and the convergence to a stationary point of that function. We may find this modification useful in both achieving faster convergence rates and in simplifying the computations involved (see the example in [30]).

(4)   At some iterations, we may leave the discrete state (signal) estimates available from the higher SNR receiver output at their current values while performing the interpolation operation using the measurements of the other receiver. This is another instance of the algorithm proposed in [30] that may simplify the computations involved and accelerate the convergence rate.

(5)   In calculating (111)-(113), we may consider substituting the full-scale smoothing by fixed-lag smoothing or even by filtering, with the obvious benefit of reduced computations and storage requirements. This is not a variant of the EM algorithm in the sense that we cannot guarantee the monotonic increase of the likelihood function or the convergence to a stationary point. However, this modification might be particularly useful in situations where the environment is changing rapidly and we want an adaptive algorithm.

Extension to Pole-Zero Processes

A state space representation of a pole-zero (ARMA) signal is given by (93)-(94), where $F(\theta)$ is given by (121) (θ are the so-called AR parameters), G is given by (122), and $h^T = (h_1...h_q)$ are the moving average (MA) parameters. The vector unknown parameters ξ to be estimated is

$$\xi = \begin{bmatrix} \tau \\ \alpha \\ \sigma_1^2 \\ \sigma_2^2 \\ h \\ \theta \end{bmatrix} \tag{160}$$

Following the considerations leading to (108)-(110), the algorithm in that case takes the form:

$$\underset{\theta}{\text{Max}}\, H^{(\ell)}(\theta) \text{======>} \hat{\theta}^{(\ell+1)} \tag{161}$$

$$\underset{\tau,\alpha,\sigma_1^2,\sigma_2^2,h}{\text{Max}}\, G^{(\ell)}(\tau,\alpha,\sigma_1^2,\sigma_2^2,h) \text{======>} \begin{matrix} \hat{\tau}^{(\ell+1)} & \hat{\alpha}^{(\ell+1)} \\ \hat{\sigma}_1^2{}^{(\ell+1)} & \hat{\sigma}_2^2{}^{(\ell+1)}, \hat{h}^{(\ell+1)} \end{matrix} \tag{162}$$

where $H^{(\ell)}(\theta)$ is given by (126), and

$$G^{(\ell)}(\tau,\alpha,\sigma_1^2,\sigma_2^2,h)$$

$$= -\frac{N}{2}\log\sigma_1^2 - \frac{1}{2\sigma_1^2}\sum_{k=1}^{N}[z_{1k}^2 - 2z_{1k}\,h^T\hat{x}^{(\ell)}(t_k) + h^T x(t_k)x(t_k)^{T^{(\ell)}}h]$$

$$- \frac{N}{2}\log\sigma_2^2 - \frac{1}{2\sigma_2^2}\sum_{k=1}^{N}[z_{2k}^2 - 2\alpha z_{2k}\,h^T\hat{x}^{(\ell)}(t_k;\tau) + \alpha^2 h^T x(t_k;\tau)x(t_k;\tau)^{T^{(\ell)}}h] \tag{163}$$

where $x(t;\tau) = x(t-\tau(t))$.

From (161)-(162), we see that the optimization with respect to (w.r.t) the AR parameters θ is decoupled from that associated with the delay τ. It is also decoupled from the optimization w.r.t the MA part of the spectral parameters. However, the latter must be performed jointly with the delay τ. Since $G^{(\ell)}(\tau,\alpha,\sigma_1^2,\sigma_2^2,h)$ in (163) is quadratic in h, the maximization w.r.t to h can easily be obtained for any specified $\tau,\alpha,\sigma_1^2,\sigma_2^2$. This suggests the following two-step procedure:

1) Up-date $r, \alpha, \sigma_1^2, \sigma_2^2$ using (114)-(117), (119)-(120) with $\mathbf{h} = \hat{\mathbf{h}}^{(\ell)}$

2) Up-date $\mathbf{h}$ via

$$\underset{\mathbf{h}}{\text{Max}} \; G^{(\ell)} (\hat{r}^{(\ell+1)}, \hat{\alpha}^{(\ell+1)}, \hat{\sigma}_1^{2(\ell+1)}, \hat{\sigma}_2^{2(\ell+1)}, \mathbf{h}) \Longrightarrow \hat{\mathbf{h}}^{(\ell+1)} \qquad (164)$$

This two-step maximization preserves the monotonic increase of the likelihood and the convergence to a stationary point of it, prehaps at some moderate reduction in speed of convergence.

The idea of decomposing the spectral optimization into separate optimizations w.r.t to the AR and the MA parts is due to Musicus and Lim [31]. In fact, this idea could also be incorporated into the algorithms presented in Section II for the stationary case.

C. Simulation Results

To demonstrate the performance of the algorithm we have considered the following example. The source radiates a continuous-time all-pole process, whose spectrum is

$$P_S(\omega) = \frac{1}{\omega^4 - 2.4375\omega^2 + 2.25}$$

The corresponding state space model of the signal is given by

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ -1.5 & -0.75 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \omega(t)$$

$$s(t) = (1 \quad 0) \, x(t)$$

The only unknown parameter to be estimated is the delay r. The data were generated using $r=3.1$ (the true delay), and sampled at a constant rate of $\Delta t = 1.0$. The noise variances are $\sigma_1^2 =$

0.25 (0.5) and σ_z^2 = 0.50 (1.0), α = 1 and N = 220 data points. Using exhaustive search, the ML estimate is found to be at $\hat{\tau}_{ML} \approx$ 3.02 (3.0). An approximate CRLB is obtained by calculating the sample variance of the score (see Eq. (167) in the sequel) using monte-carlo simulations. We find CRLB $\approx$ 0.09 (0.149). In Figure 6 we have tabulated the results, where the initial guess is $\hat{\tau}^{(0)}$ = 4.0. We observe that after few iterations, the algorithm essentially converges within the CRLB, to the ML estimate of the delay. For reference, we have also calculated the log-likelihood function $L_Z(\tau)$ along the iterations. ($L_Z(\tau)$ is computed using the Kalman filter equations).

IV. GRADIENT-BASED ALGORITHMS AND PERFORMANCE EVALUATION

As shown in [16], the rate of convergence of the algorithm (near the point of convergence) is geometrical (exponential), depending on the fraction of the complete data that can be predicted using the observed (incomplete) data. If that fraction is small (e.g., under low SNR conditions, and/or low sampling rates) the rate of convergence tends to be slow, in which case we want to use the Gaussian method or the Newton-Raphson or some other gradient search algorithm. These methods require the computation of the log-likelihood gradient (score) and the computation of the log-likelihood Hessian or the Fisher's information matrix (FIM), or some approximation of which. The FIM can further be used to assess the mean square error of the ML parameter estimates.

The switching from the EM algorithm to a gradient-based method can be facilitated using the following identity, first presented by Fisher (1925, [32]), and recently by Dempster et al. [16], Louis [33], and Meilijson [34]:

$$\frac{\partial}{\partial \xi} L_Z(\xi) = E_\xi \left\{ \frac{\partial}{\partial \xi} L_Y(\xi)/z_1, z_2, \ldots z_N \right\} =$$

$$= \frac{\partial}{\partial \xi} Q(\xi,\xi') \Big|_{\xi'=\xi} \tag{165}$$

Using (165), the observed data score can be computed by taking the conditional expectation of the complete data score. A proof of Fisher's identity when part of the complete data are continuous-time processes is presented in [35].

To simplify he exposition, we concentrate on the all-pole signal case. In view of (105) and (126),

$$\frac{\partial}{\partial \theta} L_Z(\xi) = \frac{\partial}{\partial \theta} H^{(\ell)}(\theta) \Big|_{\hat{\xi}(\ell) = \xi}$$

$$= \frac{1}{g} \left[\int_{T_i}^{T_f} \widehat{x(t)^T dx_q(t)} - \theta^T \int_{T_i}^{T_f} \widehat{x(t)x(t)^T} dt \right] \tag{166}$$

where $(\hat{\cdot}) \overset{\Delta}{=} E_\xi\{(\cdot)/z_1,z_2,...z_N\}$. Similarly, in view of (105) and (106),

$$\frac{\partial}{\partial \tau_j} L_Z(\xi) = \frac{\partial}{\partial \tau_j} G^{(\ell)}(\tau,\alpha,\sigma_1^2,\sigma_2^2) \Big|_{\hat{\xi}(\ell) = \xi}$$

$$= -\frac{1}{\sigma_2^2} \sum_{k=1}^{N} t_k^j \left[\alpha z_{2k} \widehat{s(t_k;\tau)} - \alpha^2 \widehat{s(t_k;\tau)\dot{s}(t_k;\tau)} \right] \quad j=0,1,2,... \tag{167}$$

$$\frac{\partial}{\partial \alpha} L_Z(\xi) = \frac{\partial}{\partial \alpha} G^{(\ell)}(\tau,\alpha,\sigma_1^2,\sigma_2^2) \Big|_{\hat{\xi}(\ell) = \xi}$$

$$= \frac{1}{\sigma_2^2} \sum_{k=1}^{N} [z_{2k} \widehat{s(t_k;\tau)} - \alpha \widehat{s^2(t_k;\tau)}] \tag{168}$$

$$\frac{\partial}{\partial \sigma_1^2} L_Z(\xi) = \frac{\partial}{\partial \sigma_1^2} G^{(\ell)}(\tau,\alpha,\sigma_1^2,\sigma_2^2)\Big|_{\hat{\xi}(\ell) = \xi}$$

$$= -\frac{N}{2\sigma_1^2} + \frac{1}{2\sigma_1^4} \sum_{k=1}^{N} [z_{1k}^2 - 2z_{1k}\,\widehat{s}(t_k) + \widehat{s^2(t_k)}] \tag{169}$$

$$\frac{\partial}{\partial \sigma_2^2} L_Z(\xi) = \frac{\partial}{\partial \sigma_2^2} G^{(\ell)}(\tau,\alpha,\sigma_1^2,\sigma_2^2)\Big|_{\xi(\ell) = \xi}$$

$$= -\frac{N}{2\sigma_2^2} + \frac{1}{2\sigma_2^4} \sum_{k=1}^{N} [z_{2k}^2 - 2\alpha z_{2k}\,\widehat{s}(t_k;\tau) + \alpha^2\,\widehat{s^2(t_k;\tau)}] \tag{170}$$

where we note that in the case of all-pole signals $s(t) = x_1(t)$ and $\dot{s}(t) \overset{\Delta}{=} ds(t)/dt = x_2(t)$, the second component of the $q \times 1$ state vector $x(t)$, provided that $q \geq 2$ (otherwise the process $s(t)$ is not differentiable, and the regularity conditions imposed upon Fisher's identity are violated (see [35])). With this convention, the various conditional expectations indicated in (166) - (170) are precisely those required by the EM algorithm, and they involve the continuous-discrete smoothing equations. In fact, only the discrete smoothing equations are needed for the computation of (167) - (170).

The Fisher's identity can also be applied to the stationary case. There, however, we could get at the same result by the direct differentiation of $L_Z(\xi)$ (Eq.(8)) (with the exception of stationary signals and linearly time-varying delays, where Fisher's identity allows us to differentiate first and then perform the indicated time-scale operation. By that we retain the stationarity of the underlying processes and the score computation becomes relatively easy, see the considerations in Section II-E). In the case discussed here, the direct differentiation of $L_Z(\xi)$ implies a significantly more complicated way of computing the score, although the final result should be the same. The direct differentiation of $L_Z(\xi)$ with respect to θ requires the computation of the continuous-discrete filtering equations [25] and their derivatives with respect to the components of θ (the so-called sensitivity derivatives [36]). Thus, if θ is a p-dimensional vector, the direct approach requires roughly the

equivalent of (p+1) continuous-discrete filters. Score evaluation based on the equivalent discrete model is even more complicated since it requires the differentiation of the discrete filtering equations and the associated matrices with respect to θ. The direct differentiation with respect to r may not even be expressed analytically, and one is forced to substitute the derivatives with finite differences.

Using (166) - (170) we are able to compute the score in a closed-form with a computation that only employs the Kalman smoother. We note that the continuous-discrete smoothing is performed by going through the equivalent discrete model, but there is no need to differentiate its matrices, as is needed in the direct computation of the score. We further note that the smoothed error covariance matrix can be precomputed, a feature that can be exploited for efficient computations and approximations of the score. For example, pre-computation of the error covariance matrix may indicate whether approximation can be made by performing filtering alone rather than the full smoothing with the obvious benefit of reduced computations and storage requirements.

Hessian Evaluation

The log-likelihood Hessian

$$H_Z(\xi) \stackrel{\Delta}{=} \partial^2 L_Z(\xi)/\partial \xi^2 \tag{171}$$

is computed by differentiating the expression for the score (Eqs. (166)-(170)) with respect to the components of ξ. The derivatives with respect to θ result a set of forward-backward sensitivity equations that require roughly the equivalent of (p+1) continuous-discrete Kalman smoothers. The derivatives with respect to r are not accessible since r is implicitly induced in the Kalman smoothing equations. In practice, the derivatives with respect to r may be approximated by finite differences. We perturb one coordinate of r at a time and compute the resulting score at the perturbed parameter. If r is a m-dimensional vector, this approximation requires the computation of the smoothing

equations at (m+1) closely spaced values ot τ, a task that can be simplified by pre-computation of the smoothed error covariance matrix. The same numerical procedure can be applied with respect to the components of θ. Also, pre-computation of the smoothed error covariance matrix may indicate if further approximation can be made by performing filtering alone rather than the full smoothing, resulting in a simple recursive approximation of the Hessian. Another approximation of the Hessian, based on score computations will be presented in the sequel.

FIM Evaluation

The FIM can be used to assess the mean square errors of the resulting ML parameter estimates. It can further be used in conjunction with the scoring algorithm, which is a variant of the Newton-Raphson method where the Hessian is approximated by the FIM, with a minus sign.

Assuming that the source signal and the additive noises are stationary, and that the observation interval is long compared with the correlation time (inverse bandwidth) of the signal and the noises, the FIM, computed on the basis of continuous-time observations, assumes the form [4] [22]

$$
J(\xi) \overset{\Delta}{=} E_\xi \left\{ \frac{\partial L_Z(\xi)^T}{\partial \xi} \cdot \frac{\partial L_Z(\xi)}{\partial \xi} \right\}
$$
$$
= \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix} \tag{172}
$$

where

$$
J_1 \overset{\Delta}{=} E_\xi \left\{ \frac{\partial L_Z(\xi)^T}{\partial \tau} \cdot \frac{\partial L_Z(\xi)}{\partial \tau} \right\} \tag{173}
$$

Closed form expressions for J_1 and J_2 can be found in [6] [22]. This result is an extension of the result in (15) for the case of non-stationary (moving) sources. It asserts that the quality of the delay estimates is unaffected by lack of knowledge of the spectral parameters, and vice versa.

However, this result critically depends on the stationarity of the source signal and the large time-bandwidth assumptions. For moderate observation intervals, the delay and spectral estimates are statistically correlated, the FIM is no longer block diagonal, and its computation becomes more complicated. If, in addition, we allow non-stationary signals, the direct computation of the FIM becomes exceedingly more complicated and, as far as we know, it has not been attempted yet.

By computing the second moment (covariance) of the score (Eqs. (166)-(170)), we are able to evaluate the FIM even for the non-stationary case.[1] To simplify the exposition, we concentrate on the delay estimation problem, and assume that θ, α, σ_1^2 and σ_2^2 are known.

From (167),

$$\frac{\partial}{\partial \tau_i} L_Z(\tau) = c_i - \frac{1}{\sigma_2^2} \sum_{k=1}^{N} t_k^i \left[\alpha(z_{2k} - \alpha s(t_k;\tau)) \dot{s}(t_k;\tau) \right]$$

$$= c_i - \frac{\alpha}{\sigma_2^2} \sum_{k=1}^{N} t_k^i \widehat{v_{2k} \dot{s}(t_k;\tau)} \tag{174}$$

where c_i is a constant independent of $z_1, z_2, ... z_N$.

Taking the expectation of the product $\frac{\partial}{\partial \tau_i} L_Z(\tau) \frac{\partial}{\partial \tau_j} L_Z(\tau)$ and invoking the zero-mean property of the score, that is $E_\tau \{ \frac{\partial}{\partial \tau_j} L_Z(\tau) \} = 0$ to eliminate c_i, the i,j-element of the FIM is given by

$$J_{ij}(\tau) = E_\tau \left\{ \frac{\partial}{\partial \tau_i} L_Z(\tau) \frac{\partial}{\partial \tau_j} L_Z(\tau) \right\} =$$

---

[1] We note in passing that the rationale of Fisher's identity leads also to the reversibility theorem, used by Weinstein [4] to derive the FIM for the moving source, WSS signal case.

$$= - \frac{\alpha^2}{\sigma_2{}^4} E_\tau \left\{ \sum_{m=1}^{N} t_m^i \, \hat{v}_{2m} \, \hat{s}(t_m;\tau) \, [c_j - \frac{\alpha}{\sigma_2{}^2} \sum_{\ell=1}^{N} t_\ell^j \, \hat{v}_{2\ell} \, \hat{s}(t_\ell;\tau)] \right\} \qquad (175)$$

Using the well-known formula for the expectation of the product of four zero-mean Gaussian random variables, and invoking again the zero-mean property of the score to eliminate c_j, we obtain

$$J_{ij}(\tau) = \frac{\alpha^2}{\sigma_2{}^4} \sum_{m=1}^{N} \sum_{\ell=1}^{N} t_m^i \, t_\ell^j \, [\overline{\hat{v}_{2m} \, \hat{s}(t_\ell;\tau)} \cdot \overline{\hat{v}_{2\ell} \, \hat{s}(t_m;\tau)} + \overline{\hat{v}_{2m} \, \hat{v}_{2\ell}} \cdot \overline{\hat{s}(t_\ell;\tau) \, \hat{s}(t_m;\tau)}] \qquad (176)$$

where $(\overline{\cdot}) \triangleq E_\tau(\cdot)$. Recall that for all-pole signals $s(t)=x_2(t)$, the second component of the state vector. Hence,

$$J_{ij}(\tau) = \frac{\alpha^2}{\sigma_2{}^4} \sum_{m=1}^{N} \sum_{\ell=1}^{N} t_m^i \, t_\ell^j \, [\overline{\hat{v}_{2m} \, \hat{x}_2(t_\ell;\tau)} \cdot \overline{\hat{v}_{2\ell} \, \hat{x}_2(t_m;\tau)} + \overline{\hat{v}_{2m} \, \hat{v}_{2\ell}} \cdot \overline{\hat{x}_2(t_\ell;\tau) \, \hat{x}_2(t_m;\tau)}] \qquad (177)$$

where $x_2(t;\tau) = x_2(t-\tau(t))$.

The above result corresponds to sampled data. We have therefore taken into consideration the loss of information due to aliasing effects.

Observing that $v_{2k} = z_{2k} - \alpha h^T x(t_k;\tau)$, the elements of the FIM are expressible in terms of $E_\tau\{\hat{x}(t_m;\tau)\hat{x}(t_\ell;\tau)^T\} \triangleq P_{n;\ell}$ the covariance matrix of the smoothed estimates of the state.

The expression for the FIM involves the double sum over N, the number of data points, and therefore tends to be computationally expensive. Successive approximat ions of the FIM, based on substituting the full-scale smoothed estimates $\hat{x}_2(t_k;\tau)$ and $\hat{v}_{2k}$ with fixed-lag smoothing of increasing order, can be obtained following the considerations in [37].

An alternative approximation of the FIM can be obtained by decomposing the score into conditional scores

$$\frac{\partial L_Z(\xi)}{\partial \xi} = \sum_{r=1}^{N} \frac{\partial L_{r|r-1}(\xi)}{\partial \xi} \tag{178}$$

where

$$L_{r|r-1}(\xi) = \begin{array}{ll} \log f(z_1;\xi) & r=1 \\[8pt] \log f(z_r/z_{r-1},...z_1;\xi) & r=2,3,...N \end{array} \tag{179}$$

where $f(\cdot)$ stands for the appropriate probability density function and the z_r's are the scaler measurements defined in Eq. (134). Eq. (178) is a martingale representation of the score since the increments are the conditional scores, and therefore have zero conditional expectation given the past. Following [34], we propose the following approximation of the FIM:

$$\hat{J}(\xi) = \sum_{r=1}^{N} \left[\frac{\partial}{\partial \xi} L_{r|r-1}(\xi) - \frac{1}{N} \sum_{\ell=1}^{N} \frac{\partial}{\partial \xi} L_{\ell|\ell-1}(\xi) \right]^T \left[\frac{\partial}{\partial \xi} L_{r|r-1}(\xi) - \frac{1}{N} \sum_{\ell=1}^{N} \frac{\partial}{\partial \xi} L_{\ell|\ell-1}(\xi) \right]$$

$$= \sum_{r=1}^{N} \frac{\partial}{\partial \xi} L_{r|r-1}(\xi)^T \frac{\partial}{\partial \xi} L_{r|r-1}(\xi) - \frac{1}{N} \frac{\partial}{\partial \xi} L_Z(\xi)^T \frac{\partial}{\partial \xi} L_Z(\xi) \tag{180}$$

This estimate is, in fact, the sample covariance matrix of the conditional scores, where the actual FIM is the covariance matrix of the score. The estimate may also be used, with a minus sign, to approximate the Hessian. Since the increments of a martingale are uncorrelated, the expected value of the sum of squares of the conditional scores (the first term on the second line of (180)) is an unbaised estimate of the covariance matrix of their sum, which is the FIM. The expected value of the second term of that expression equals the FIM divided by N. Therefore, $\hat{J}(\xi)/(N-1)$ has the

same expectation as FIM/N (i.e., "asymptotic unbiasedness"). Furthermore, consistency of the estimate can be shown under stationarity and ergodicity of the martingale-difference increments (i.e., the conditional scores), e.g., see [38] [39]. We note that under the usual regularity conditions for ML to apply, the consistency will be preserved if we replace ξ by its ML estimate.

Now, since

$$f(z_r/z_{r-1},...z_1;\xi) = \frac{f(z_r,...z_1;\xi)}{f(z_{r-1},...z_1;\xi)} \tag{181}$$

it immediately follows that

$$\frac{\partial}{\partial\xi} L_{r|r-1}(\xi) = \frac{\partial}{\partial\xi} L_r(\xi) - \frac{\partial}{\partial\xi} L_{r-1}(\xi) \tag{182}$$

where $L_r(\xi)$ denotes the log-likelihood function of the observed data $z_1,...z_r$. Substituting (182) into (180), the proposed estimate can be computed using (166)-(170).

V. CONCLUSIONS

We have developed computationally efficient iterative algorithms for finding the Maximum Likelihood estimates of the delay and spectral parameters of a noise-like Gaussian signal radiated from a common point source and observed by two or more spatially separated receivers. We first consider the stationary case in which the source is stationary (not moving) and the observed signals are modeled as wide sense stationary processes. We then extend the scope by considering a non-stationary (moving) source radiating a possible non-stationary stochastic signal. In that context, we address the practical problem of estimation given discrete-time observations. We also present efficient methods for calculating the log-likelihood gradient (score), the Hessian, the Fisher's information matrix under stationary and non-stationary conditions.

Acknowledgements

References

[1]   Special Issue on Time Delay Estimation, IEEE Trans. Acoust., Speech, and Signal Processing, June 1981.

[2]   V.H. MacDonald and P.M. Schultheiss, "Optimum passive bearing estimation in a spatially incoherent noise environment", J. Acoust. Soc. Amer., Vol. 46, No. 1 (Pt. 1), pp. 37-43, 1969.

[3]   C.H. Knapp and G.C. Carter, "The generalized correlation method for estimation of time delay", IEEE Trans. Acoust., Speech, and Signal Processing, Vol. ASSP-24, No. 4, pp. 320-327, August 1976.

[4]   E. Weinstein, "Estimation of trajectory parameters from passive array measurements", Ph.D. dissertation, Yale Univ., New Haven, CT, 1978.

[5]   C.H. Knapp and G.C. Carter, "Estimation of time delay in the presence of source or receiver motion", J. Acoust. Soc. Amer., Vol. 61, No. 6, pp.1545-1549, June 1977.

[6]   E. Weinstein and D. Kletter, "Delay and Doppler estimation by time-space partition of the array data", IEEE Trans. Acoust., Speech, and Signal Processing, Vol. ASSP-31, No. 6, pp. 1523-1535, Dec. 1983.

[7]   J.A. Stuller, "Maximum likelihood estimation of time-varying delay - Part I", IEEE Trans. Acoust., Speech, and Signal Processing, Vol. ASSP-35, No. 3, pp. 300-313, March 1987.

[8]   M. Namazi and J.A. Stuller, "A new approach to signal registration with the emphasis on variable time delay estimation", IEEE Trans. Acoust., Speech, and Signal Processing, Vol. ASSP-35, No.12, pp. 1649-1660, Dec. 1987.

[9]   Y.T. Chan, J.M. Riley, and J.B. Plant, "A parameter estimation approach to time delay estimation and signal detection", IEEE Trans. Acoust., Speech, and Signal Processing, Vol. ASSP-28, No. 1, pp. 8-16, Feb. 1980.

[10]   E.J. Hannan and P.J. Thomson, "Estimating group delay", Biometrika, Vol. 60, pp. 241-253, 1973.

[11]   B.V. Hamon and E.J. Hannan, "Spectral estimation of time delay for dispersive and non-dispersive systems", J. Roy. Stat. Soc. Ser. C. (Appl. Statist.), Vol. 23, No. 2, pp. 134-142, 1974.

[12]   G.C. Carter, A.H. Nuttall, and P.G. Cable, "The smoothed coherence transform", Proc. IEEE (Lett.), Vol. 61, pp. 1497-1498, Oct. 1973.

[13]   P.R. Roth, "Effective measurements using digital signal analysis", IEEE Spectrum, Vol. 8, pp. 62-70, Apr. 1971.

[14]   E.J. Hannan and P.J. Thomson, "Delay estimation and the estimation of coherence and phase", IEEE Trans. Acoust., Speech, and Signal Processing, Vol. ASSP-29, No. 3, pp. 485-490, June 1981.

[15]   H.L. Van Trees, Detection, Estimation and Modulation Theory, Part I. New York: Wiley, 1968.

[16]   A.P. Dempster, N M. Laird and D.B. Rubin, "Maximum likelihood estimation from incomplete data via the EM algorithm", J. Roy. Stat. Soc. Ser. B., Vol. 39, pp. 1-38, 1977.

[17]   C.F.J. Wu, "On the convergence properties of the EM algorithm", The Annals of Statistics, No. 11, pp. 95-103, 1983.

[18]   B.D.O. Anderson and J.B. Moore, Optimal Filtering. New Jersey: Prentice Hall, Inc., 1979.

[19]   W.J. Bangs and P.M. Schultheiss, "Space-time processing for optimal parameter estimation", in Signal Processing (Proc. NATO Advanced Study Institute, Loughborough, England), J.W.R. Griffith, P.L. Stocklin and C. Van Schooneveld, Eds. New York: Academic, 1973.

[20]   G.C. Carter, "Variance bounds for passively locating an acoustic source with a symmetric line array", J. Acoust. Soc. Amer., Vol. 62, No.4, pp. 922-926, Oct. 1977.

[21]   W.R. Hann, "Optimum signal processing for passive sonar range and bearing estimation", J. Acoust. Soc. Amer., Vol.58, No.2, pp. 201-207, 1975.

[22]   P.M. Schultheiss and E. Weinstein, "Lower bounds on the localization errors of a moving source observed by a passive array", IEEE Trans. Acoust., Speech, and Signal Processing, Vol. ASSP-29, No. 3, pp. 600-607, June 1981.

[23] M. Segal and E. Weinstein, "Maximum likelihood parameter identification of continuous-time linear dynamic systems given discrete-time observations", Proceeding of the IFAC Conf. Identification '88, Beijing, China, Aug. 1988.

[24] A. Le Breton, "On continuous and discrete sampling for parameter estimation in diffusion type processes", in Mathematical Programming Studies, 5, pp. 124-144, 1976.

[25] A.H. Jazwinski, Stochastic Processes and Filtering Theory. New York: Academic Press, 1970.

[26] M. Segal, "Optimal parameter estimation from incomplete data and its application to time-space signal processing", Ph.D. dissertation, Tel Aviv Univ., 1988.

[27] I.V. Basawa and B.L.S. Prakasa Rao, Statistical Inference for Stochastic Processes. London: Academic Press, 1980.

[28] P.S. Maybeck, Stochastic Models, Estimation and Control, Vol. 1. Orlando: Academic Press, 1979.

[29] R.H. Shumway, "Some applications of the EM algorithm to analyzing incomplete time series data", in Time Series Analysis of Irregularly Observed Data", E. Parzen, Ed., Lecture Notes in Statistics. New York: Springer-Verlag, 1984.

[30] M. Segal and E. Weinstein, "The Cascade EM-Algorithm", accepted for publication, Proc. IEEE (Lett.)

[31] B.R. Musicus and J.S. Lim, "Maximum likelihood parameter estimation of noisy data", IEEE Int. Conf. Acoust., Speech, and Signal Processing, Washington, D.C., pp. 224-227, April 1979.

[32] R.A. Fisher, "Theory of statistical estimation", Proc. Cambridge Phil. Soc., Vol. 22, pp. 700-725, 1925.

[33] T.A. Louis, "Finding the observed information matrix when using the EM algorithm", J. Roy. Soc. Ser. B., Vol. 44, No. 2, pp. 226-233, 1982.

[34] I. Meilijson, "A fast improvement to the EM algorithm on its own terms", to be published in J. Roy. Stat. Soc. Ser. B.

[35]   M. Segal and E. Weinstein, "A new method for evaluating the log-likelihood gradient (score) of linear dynamic systems", to be published in the IEEE Trans. Auto. Control.

[36]   N.R. Sandell, Jr. and K.I. Yared, "Maximum likelihood identification of state space models for linear dynamic systems", MIT Tech. Rep. ESL-R-814, Apr. 1978.

[37]   M. Segal and E. Weinstein, "A new method for evaluating the log-likelihood gradient, the Hessian and the Fisher's infomation matrix of linear dynamic systems", accepted for publication (IEEE Info. Th.).

[38]   P. Billingsley, Convergence of Probability Measures. New York: Wiley, 1968.

[39]   P. Holewijn and I. Meilijson, "Notes on the central limit theorem for stationary sequences", Sem. Prob. Strasb., XVII, L.N. 986, Springer, pp. 240-242, 1981/2.

Figure Captions

Figure 1:   Two-channel time delay estimator - the stationary case.

Figure 2:   Simulation results.

Figure 3:   The multi-receiver vector delay estimator.

Figure 4:   Two-channel time delay and Doppler estimator.

Figure 5:   Two-channel time delay estimator – the non-stationary case.
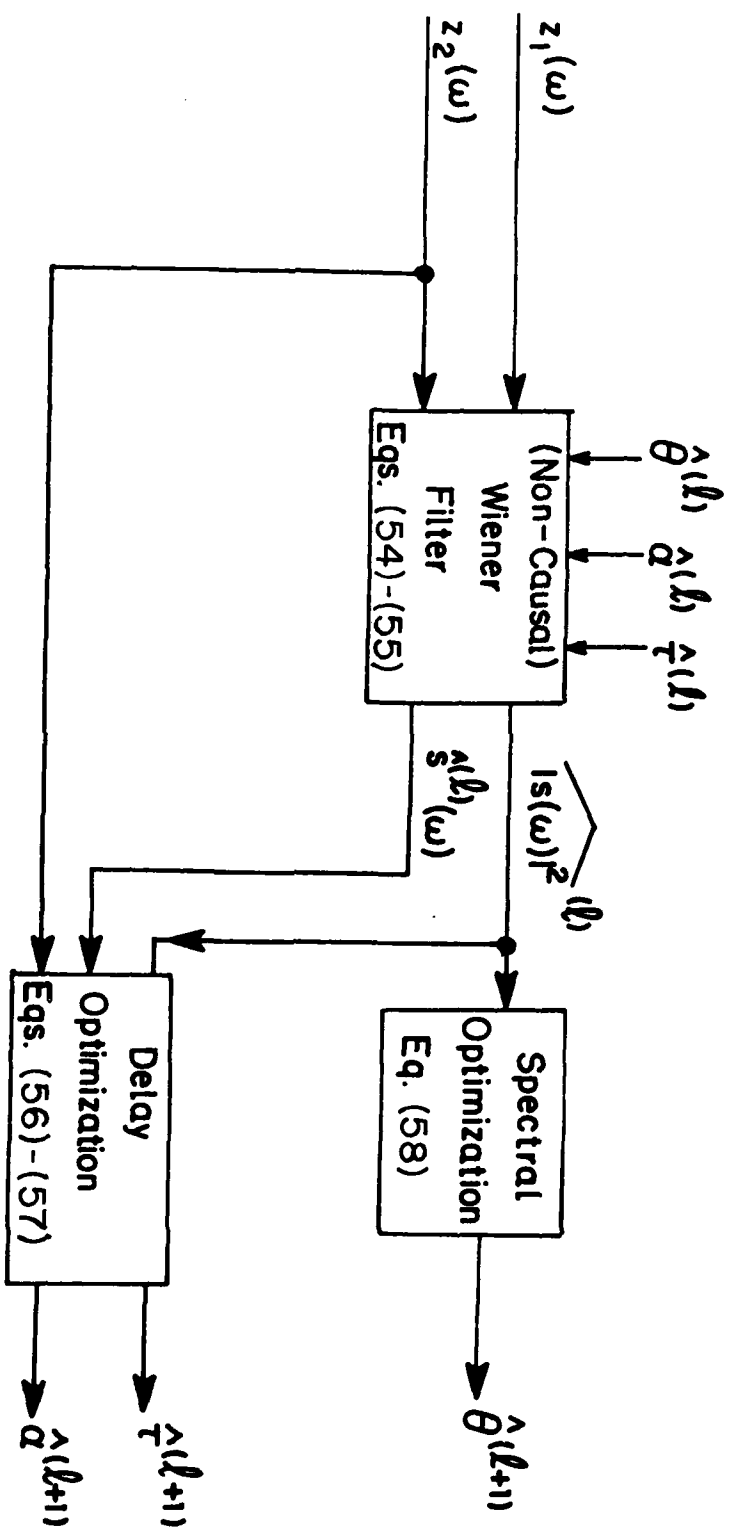
Figure 6:   Simulation results.

Figure 1 : TWO-CHANNEL TIME DELAY ESTIMATOR-THE STATIONARY CASE

$\hat{\tau}^{(0)}/\Delta t = 10.0, \quad \hat{\alpha}^{(0)} = .7, \quad \hat{\theta}_1^{(0)} = -1.8, \quad \hat{\theta}_2^{(0)} = 1.7, \quad \hat{\theta}_3^{(0)} = -.8$

ℓ	$\hat{\tau}^{(\ell)}/\Delta t$	$\hat{\alpha}^{(\ell)}$	$\hat{\theta}_1^{(\ell)}$	$\hat{\theta}_2^{(\ell)}$	$\hat{\theta}_3^{(\ell)}$	$L_Z(\hat{\xi}^{(\ell)})/N$
1	10.22	.912	-1.813	1.647	-.748	-4.1870
2	10.33	.992	-1.806	1.614	-.714	-4.1256
3	10.39	1.031	-1.796	1.591	-.695	-4.1144
4	10.43	1.052	-1.787	1.575	-.684	-4.1114
5	10.46	1.065	-1.781	1.564	-.678	-4.1104
6	10.48	1.073	-1.776	1.557	-.674	-4.1100
7	10.49	1.078	-1.774	1.554	-.673	-4.1098

TABLE 1

Figure 2 : SIMULATION RESULTS

$\hat{\tau}^{(0)}/\Delta t = 11.0, \quad \hat{\alpha}^{(0)} = .7, \quad \hat{\theta}_1^{(0)} = -1.8, \quad \hat{\theta}_2^{(0)} = 1.7, \quad \hat{\theta}_3^{(0)} = -.8$

ℓ	$\hat{\tau}^{(\ell)}/\Delta t$	$\hat{\alpha}^{(\ell)}$	$\hat{\theta}_1^{(\ell)}$	$\hat{\theta}_2^{(\ell)}$	$\hat{\theta}_3^{(\ell)}$	$L_Z(\hat{\xi}^{(\ell)})/N$
1	10.79	.913	-1.810	1.644	-.748	-4.1888
2	10.69	.994	-1.803	1.610	-.713	-4.1273
3	10.63	1.033	-1.793	1.587	-.694	-4.1158
4	10.59	1.054	-1.785	1.572	-.683	-4.1123
5	10.57	1.067	-1.779	1.562	-.677	-4.1109
6	10.55	1.075	-1.777	1.559	-.676	-4.1105
7	10.54	1.079	-1.775	1.557	-.675	-4.1102

TABLE 2

Figure 2 : SIMULATION RESULTS

$\hat{r}(0)/\Delta t = 10.0, \quad \hat{\alpha}(0) = .7, \quad \hat{\theta}_1(0) = -2.2, \quad \hat{\theta}_2(0) = 2.2, \quad \hat{\theta}_3(0) = -1.5$

ℓ	$\hat{r}(\ell)/\Delta t$	$\hat{\alpha}(\ell)$	$\hat{\theta}_1(\ell)$	$\theta_2(\ell)$	$\hat{\theta}_3(\ell)$	$L_Z(\hat{\xi}(\ell))/N$
1	10.215	1.328	-2.094	2.234	-1.375	-4.4730
2	10.295	1.399	-1.930	2.190	-1.130	-4.1854
3	10.35	1.342	-1.979	2.224	-1.119	-4.1334
4	10.395	1.289	-2.012	2.244	-1.112	-4.1238
5	10.425	1.246	-2.032	2.256	-1.107	-4.1184
6	10.45	1.213	-2.046	2.262	-1.104	-4.1155
7	10.465	1.189	-2.055	2.266	-1.101	-4.1139

TABLE 3

Figure 2: SIMULATION RESULTS

$\hat{\tau}^{(0)}/\Delta t = 11.0$, $\hat{\alpha}^{(0)} = .7$, $\hat{\theta}_1^{(0)} = -2.2$, $\hat{\theta}_2^{(0)} = 2.2$, $\hat{\theta}_3^{(0)} = -1.5$

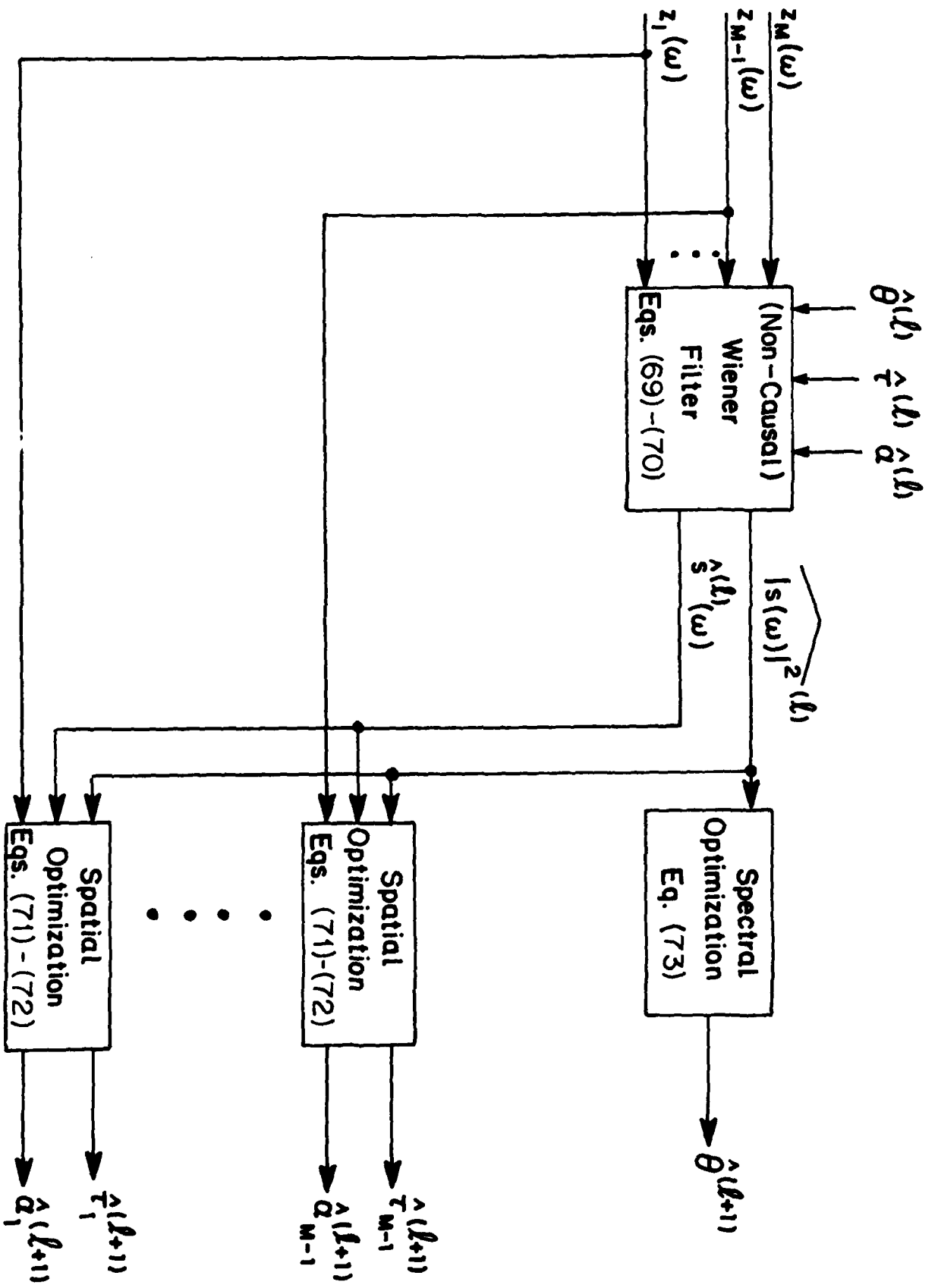| ℓ	$\hat{\tau}^{(\ell)}/\Delta t$	$\hat{\alpha}^{(\ell)}$	$\hat{\theta}^{(\ell)}$	$\hat{\theta}_2^{(\ell)}$	$\hat{\theta}_3^{(\ell)}$	$L_Z(\hat{\xi}^{(\ell)})/N$
1	10.80	1.322	-2.092	2.230	-1.373	-4.4721
2	10.72	1.396	-1.926	2.185	-1.129	-4.1875
3	10.655	1.341	-1.976	2.220	-1.118	-4.1347
4	10.62	1.289	-2.009	2.241	-1.112	-4.1250
5	10.585	1.246	-2.030	2.254	-1.107	-4.1193
6	10.56	1.214	-2.044	2.261	-1.104	-4.1161
7	10.54	1.190	-2.053	2.265	-1.101	-4.1144

TABLE 4

Figure 2: SIMULATION RESULTS

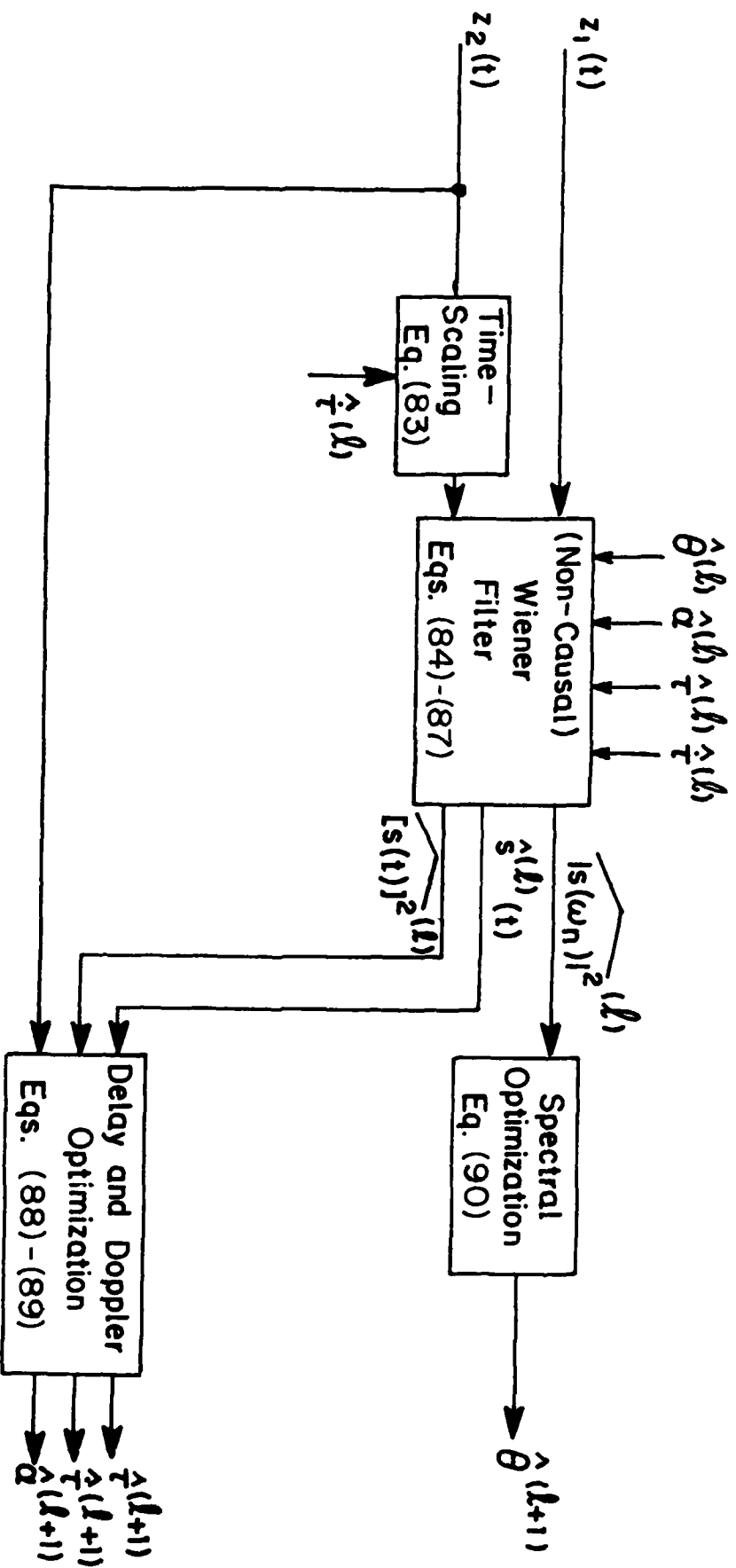Figure 3 : THE MULTI-RECEIVER VECTOR DELAY ESTIMATOR
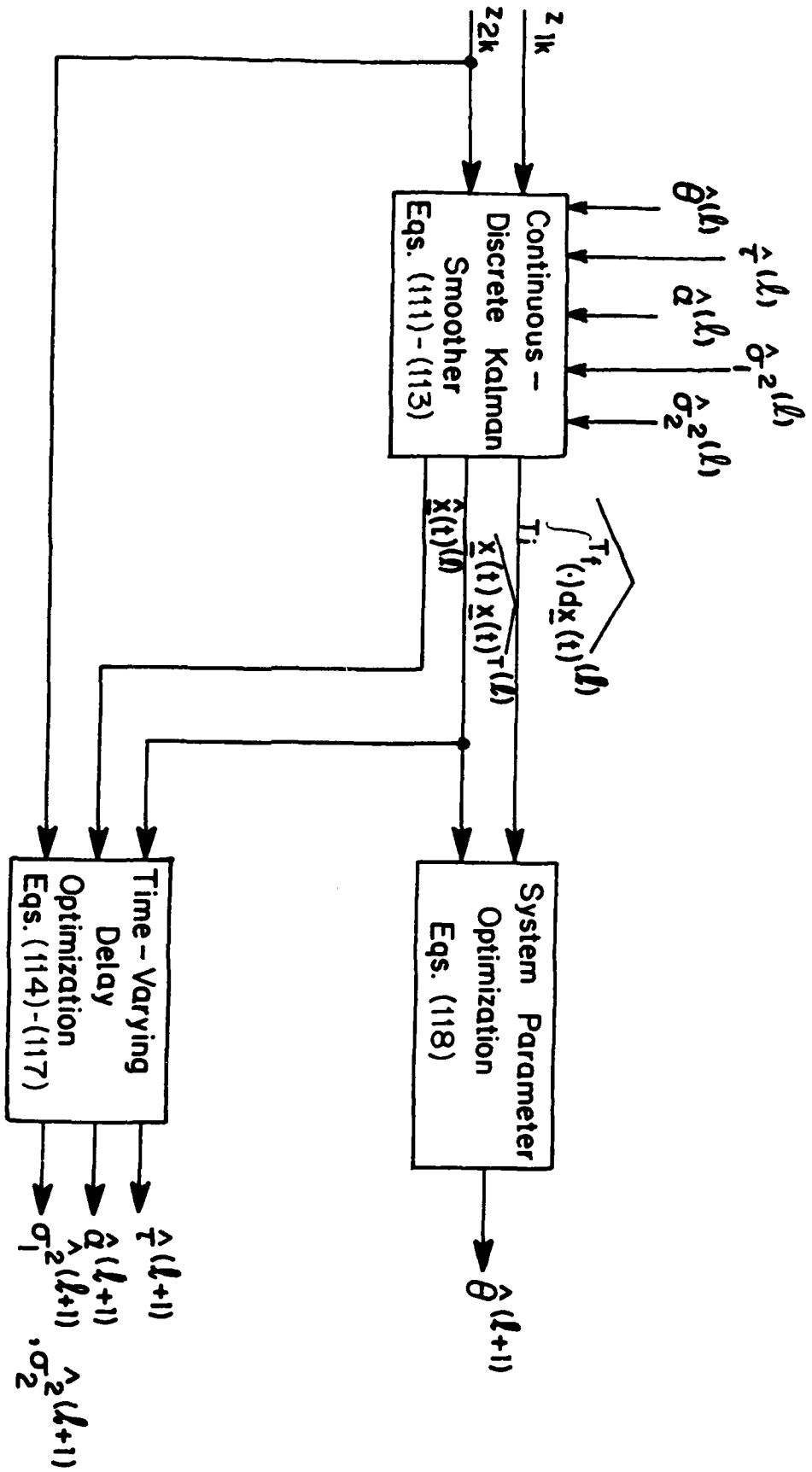
Figure 4 : TWO-CHANNEL TIME DELAY AND DOPPLER ESTIMATOR

Figure 5 : TWO-CHANNEL TIME DELAY ESTIMATOR - THE NON-STATIONARY CASE

$\sigma_1{}^2 = 0.25 \quad \sigma_2{}^2 = 0.5$

ℓ	$\hat{\tau}(\ell)$	$L_Z(\hat{\tau}^{(\ell)})$
1	3.-	-156.846
2	3.01	- 98.150
3	3.014	- 98.128
4	3.016	- 98.123

TABLE 5

Figure 6 : SIMULATION RESULTS

$\sigma_1{}^2 = 0.5 \quad \sigma_2{}^2 = 1.0$

ℓ	$\hat{r}(\ell)$	$L_Z(\hat{r}(\ell))$
1	3.6	-220.239
2	3.4	-215.162
3	3.3	-213.412
4	3.2	-212.214
5	3.1	-211.578
6	3.08	-211.518
7	3.07	-211.497
8	3.061	-211.482

TABLE 6

Figure 6: SIMULATION RESULTS

DOCUMENT LIBRARY

August 9, 1988

Distribution List for Technical Report Exchange

Attn: Stella Sanchez-Wade
Documents Section
Scripps Institution of Oceanography
Library, Mail Code C-075C
La Jolla, CA 92093

Hancock Library of Biology &
   Oceanography
Alan Hancock Laboratory
University of Southern California
University Park
Los Angeles, CA 90089-0371

Gifts & Exchanges
Library
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, NS, B2Y 4A2, CANADA

Office of the International
   Ice Patrol
c/o Coast Guard R & D Center
Avery Point
Groton, CT 06340

Library
Physical Oceanographic Laboratory
Nova University
8000 N. Ocean Drive
Dania, FL 33304

NOAA/EDIS Miami Library Center
4301 Rickenbacker Causeway
Miami, FL 33149

Library
Skidaway Institute of Oceanography
P.O. Box 13687
Savannah, GA 31416

Institute of Geophysics
University of Hawaii
Library Room 252
2525 Correa Road
Honolulu, HI 96822

Library
Chesapeake Bay Institute
4800 Atwell Road
Shady Side, MD 20876

MIT Libraries
Serial Journal Room 14E-210
Cambridge, MA 02139

Director, Ralph M. Parsons Laboratory
Room 48-311
MIT
Cambridge, MA 02139

Marine Resources Information Center
Building E38-320
MIT
Cambridge, MA 02139

Library
Lamont-Doherty Geological
   Observatory
Colombia University
Palisades, NY 10964

Library
Serials Department
Oregon State University
Corvallis, OR 97331

Pell Marine Science Library
University of Rhode Island
Narragansett Bay Campus
Narragansett, RI 02882

Working Collection
Texas A&M University
Dept. of Oceanography
College Station, TX 77843

Library
Virginia Institute of Marine Science
Gloucester Point, VA 23062

Fisheries-Oceanography Library
151 Oceanography Teaching Bldg.
University of Washington
Seattle, WA 98195

Library
R.S.M.A.S.
University of Miami
4600 Rickenbacker Causeway
Miami, FL 33149

Maury Oceanographic Library
Naval Oceanographic Office
Bay St. Louis
NSTL, MS 39522-5001

Marine Sciences Collection
Mayaguez Campus Library
University of Puerto Rico
Mayagues, Puerto Rico 00708

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. WHOI-88-29	2.	3. Recipient's Accessi n No.
4. Title and Subtitle Time Delay Estimation in Stationary and Non-Stationary Environments			5. Report Date July 1988
			6.
7. Author(s) Mordechai Segal and Ehud Weinstein			8. Performing Organization Rept. No. WHOI-88-29
9. Performing Organization Name and Address The Woods Hole Oceanographic Institution Woods Hole, Massachusetts 02543			10. Project/Task/Work Unit No.
			11. Contract(C) or Grant(G) No. (C) N00014-85-K-0272 (G)
12. Sponsoring Organization Name and Address Naval Air Systems Command			13. Type of Report & Period Covered Technical Report
			14.

15. Supplementary Notes

This report should be cited as: Woods Hole Oceanog. Inst. Tech. Rept., WHOI-88-29.

16. Abstract (Limit: 200 words)

We develop computationally efficient iterative algorithms for finding the Maximum Likelihood estimates of the delay and spectral parameters of a noise-like Gaussian signal radiated from a common point source and observed by two or more spatially separated receivers. We first consider the stationary case in which the source is stationary (not moving) and the observed signals are modeled as wide sense stationary processes. We then extend the scope by considering a non-stationary (moving source) radiating a possibly non-stationary stochastic signal. In that context, we address the practical problem of estimation given discrete-time observations. We also present efficient methods for calculating the log-likelihood gradient (score), the Hessian, and the Fisher's information matrix under stationary and non-stationary conditions.

17. Document Analysis   a. Descriptors

1. Array Processing
2. Multiple Source
3. EM Algorithm

b. Identifiers/Open-Ended Terms

c. COSATI Field/Group

18. Availability Statement Approved for publication; distribution unlimited.	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 72
	20. Security Class (This Page)	22. Price

(See ANSI-Z39.18)   See Instructions on Reverse   OPTIONAL FORM 272 (4-77)
                                                                                    (Formerly NTIS-35)
                                                                                    Department of Commerce